# An investigation of the epigenetic and transcriptional changes which follow Epstein-Barr Virus infection of germinal centre B cells

**by**

**Sarah Miriam Leonard**

**A thesis submitted to**

**The University of Birmingham**

**for the degree of**

**DOCTOR OF PHILOSOPHY**

**School of Cancer Sciences**

**The University of Birmingham**

**August 2010**

## Abstract

Although Epstein–Barr virus (EBV) usually establishes a harmless infection in human memory B cells, it is implicated in the development of germinal centre (GC) B-cell-derived malignancies, including Hodgkin's lymphoma (HL). I have shown using gene expression profiling that lymphoblastoid cell lines derived from GC B cells are a useful model for studying early EBV-associated changes contributing to the pathogenesis of HL. EBV infection of GC B cells is followed by the up-regulation of the DNA methyltransferases, DNMT3A, and the down-regulation of DNMT1 and DNMT3B, a pattern of expression which is re-capitulated in HL cell lines. I have also shown that the major EBV oncogene, LMP1, is responsible for the down-regulation in GC B cells of DNMT1, and that DNMT3A binds to the EBV promoter, Wp which is silenced by DNA methylation. Genome-wide promoter arrays revealed that EBV infection of GC B cells is followed by methylation changes in a substantial number of cellular genes. These changes were not randomly distributed across the genome but clustered at certain chromosomal locations and were strongly associated with the CpG content of gene promoters. Finally, I have shown that EBV also modulates the expression of another set of epigenetic regulators which control arginine methylation.

*Do mo theaghlach agus mo cháirde*

## Acknowledgements

I would like to thank my supervisors Professor Ciaran Woodman, Professor Paul Murray and Dr. Wenbin Wei for all their time, patience and expert advice over the years. I would like to thank my family, Brian, Dolores and Rebecca, for their support and encouragement, and for providing me with the opportunities that have taken me to where I am today. Thank you so much to all the Murray and Woodman group, past and present, especially to the girlies who helped make it such a fun and great place to work. I would also like to thank my brilliant Brummie and Irish friends for their support and friendship over the years.

# Contents

*Title of Section* *Page Number*

## Chapter 3. A genome wide analysis of transcriptional and methylation changes observed following Epstein Barr Virus infection of Germinal Centre B cells

## Chapter 5. The impact of EBV induced changes in the expression of DNA methyltransferases on the methylation and expression of viral genesHodgkin's lymphoma cells

# List of Figures

# List of Tables

# List of common abbreviations

| | |
|---|---|
| BART | *Bam* HIA rightward transcript |
| BCR | B-cell receptor |
| BL | Burkitt's lymphoma |
| bp | base pair |
| cDNA | complementary-DNA |
| cHL | classical HL |
| Cp | *Bam* HI C fragment promoter |
| DNA | deoxyribonucleic acid |
| DNMT | dna methyltransferase |
| EBER | Epstein-Barr encoded RNA |
| EBNA | Epstein-Barr nuclear antigen |
| EBNA-LP | EBNA leader protein |
| EBV | Epstein-Barr virus |
| ESC | Embryonic stem cell |
| FCS | fetal calf serum |
| g | gram |
| GAPDH | glyseraldehyde-3-phosphate dehydrogenase |
| GC | germinal centre |
| H3 | histone protein 3 |
| HIV | human immunodeficiency virus |
| HL | Hodgkin's lymphoma |
| HPV | Human papilloma virus |
| HRS | Hodgkin and Reed-Sternberg |
| ICF | Immunodeficiency, centromere instability and facial anomalies syndrome |
| Ig | immunoglobulin |
| IGB | integrated genome browser |
| IM | infectious mononucleosis |
| Kbp | kilobasepairs |
| KDa | kilodalton |
| KSHV | Kaposi's sarcoma-associated herpesvirus |
| LANA | latency-associated nuclear protein |
| L | litre |
| L&H | lymphocytic and histiocytic |
| LCL | lymphoblastoid cell line |
| LCV | lymphocryptovirus |
| LMP | latent membrane protein |
| LOH | loss of heterozygosity |
| MHC | major histocompatibility complex |
| MOI | multiplicity of infection |
| miRNA | microRNAs |
| µl | microlitre |
| mA | miliamps |
| ml | milliliter |
| mRNA | messenger-RNA |
| M | molar |

| | |
|---|---|
| ng | nanogram |
| NHL | non-Hodgkin's lymphoma |
| nLPHL | nodular lymphocyte predominant HL |
| NPC | nasopharyngeal carcinoma |
| NS | nodular sclerosis |
| OriP | origin of plasmid replication |
| PBS | phosphate-buffered saline |
| PCR | polymerase chain reaction |
| PRMT | protein arginine methyltransferase |
| PTLD | post-transplant lymphoproliferative disease |
| Qp | *Bam* HI Q fragment promoter |
| RNA | ribonucleic acid |
| Rpm | rotation per minute |
| RT-PCR | reverse transcriptase polymerise chain reaction |
| SAM | S-adenosyl-methionine |
| SAM | significance analysis of microarrays |
| TMC | tonsillar mononuclear cells |
| TF | transcription factor |
| TR | terminal repeat |
| TSG | tumour suppressor gene |
| V | volts |
| WHO | World Health Organization |
| Wp | *Bam* HI W fragment promoter |
| X-ChIP | cross-linked chromatin immunoprecipitation |
| ZRE | Zta response elements |

Chapter 1

# Introduction

## 1.1 Introduction

The introduction to my thesis is in three sections. The first section provides a brief introduction to epigenetics before focusing on the DNA methyltransferases, and the relationship between DNA methylation and transcriptional changes, and how these changes contribute to carcinogenesis. I also introduce the protein arginine methyltransferases (PRMTs). In the second section, I introduce the Epstein-Barr virus (EBV) before outlining how the expression of its genes varies during the viral life-cycle, and how the expression of these genes might be epigenetically regulated. Finally in the third section, I briefly review how oncogenic viruses disrupt the cellular epigenome before focusing on how EBV-induced methylation changes in cellular genes contribute to the pathogenesis of EBV-associated malignancies.

## 1.2 Overview of Epigenetics

Epigenetics is the study of mitotically heritable changes in gene expression not encoded by the DNA sequence. Epigenetic modifications are essential for normal development and the maintenance of gene expression patterns in mammalian cells. A failure to maintain heritable epigenetic marks can result in inappropriate activation or inhibition of signalling pathways, leading to disease (Robertson 2005).

Epigenetic modifications are dependent upon some change in chromatin structure which defines how genetic information is organised within a cell. Chromatin is made up of nucleosomes which consist of ~146 base pairs of DNA wrapped around an octamer of four core histone proteins (H3, H4, H2A and H2B) (Luger *et al.,* 1997). The epigenetic processes which modify chromatin structure include: DNA methylation, covalent histone modifications, non-covalent mechanisms (nucleosome remodelling) and microRNAs. Co-operation between

these processes regulates the accessibility and compactness of chromatin, and in so doing modulate gene expression.

### 1.2.1 DNA methylation

Methylation is the only known epigenetic modification of DNA. Methylation is carried out by DNA methyltransferases (DNMTs) using S-adenosyl-methionine (SAM) as the methyl donor (Herman and Baylin 2003) (Figure 1.1). In mammals, methylation primarily occurs at the carbon-5 position of cytosine residues within CpG dinucleotides. CpG dinucleotides are not evenly distributed across the genome but are instead found clustered in CpG rich DNA stretches called CpG islands and in repetitive sequences such as centromeric repeats, retrotransposon elements and rDNA. CpG islands are defined as regions 500 bp in length with a G+C content greater than 55% (Takai and Jones 2002). There are around 30000 CpG islands in the human genome; these are generally associated with the 5' end of genes and are found in ~60% of human gene promoters and first exons. Whereas the majority of CpG sites throughout the genome are methylated, CpG islands remain unmethylated in the majority of promoters (Illington and Bird 2009).

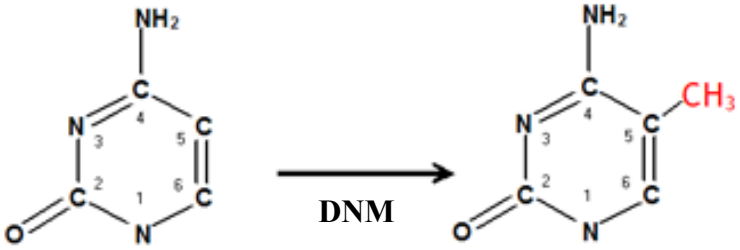



**Figure 1.1: DNA cytosine methylation at ring carbon C5.** Methylation of cytosine within CpG dinucleotides is conducted by the methyltransferases that use S-adenosyl-L-methionine as a donor of a methyl group (Diagram adapted from Herman and Baylin 2003).

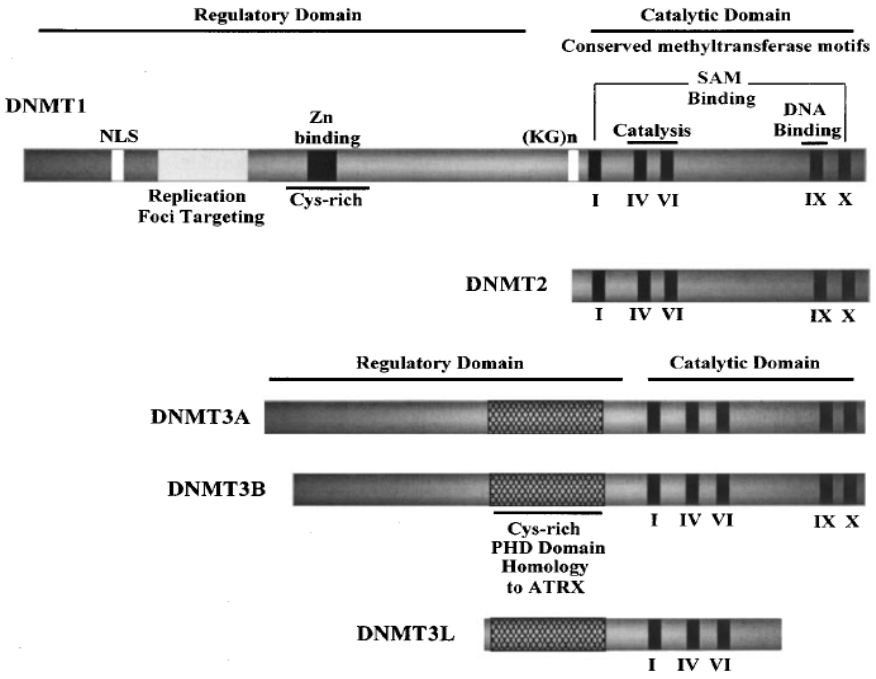



**Figure 1.2: Structure of the known DNMT proteins.** DNMT1, DNMT3A and DNMT3B can be divided into two domains, regulatory and catalytic. Conserved motifs (roman numerals) involved in catalysis are indicated with black boxes. Other structural features such as the replication foci targeting domain and zinc binding region are also indicated (Diagram from Robertson 2001).

### 1.2.2 DNA methylation in normal cells

DNA methylation is essential for the normal function and development of mammalian cells. It is a stable modification that is inherited through cell divisions allowing daughter cells to retain the same expression patterns as the parent cell, a process that is particularly important for genomic imprinting and X chromosome inactivation in females (Bird 2002). In somatic cells, changes in the pattern of DNA methylation first contribute to embryonic development, and then at a later stage to lineage specificity and terminal differentiation (Santos *et al.,* 2002). DNA methylation also ensures silencing of transposons and other parasitic elements that have been acquired by the genome over time, and which were they to be activated, would be detrimental to the cell (Reik 2007).

The distribution of methylation changes in normal cell backgrounds is now being investigated using genome wide methylation arrays (Rauch *et al.,* 2009; Yang *et al.,* 2010). A global explanation as to why some genes are methylation targets, whereas others are not, has yet to be provided. This is one of the issues addressed in my thesis in the context of germinal centre (GC) B cells infected with EBV.

## 1.3 DNA methyltransferases

The mammalian DNA methyltransferase (DNMT) family which comprises DNMT1, DNMT3A and DNMT3B together with the accessory protein DNMT3L are responsible for the regulation of DNA methylation in all cell types (Figure 1.2). The DNMTs can be divided into two groups, the maintenance methylation enzyme (DNMT1), and the *de novo* methylation enzymes (DNMT3A and DNMT3B) (Robertson 2001). DNMT2, a protein that was originally identified as a DNMT has a highly similar sequence and structure to that of the other DNMTs and is responsible for the methylation of cytosine 38 in the anticodon loop of $tRNA^{ASP}$; however, it has little or no DNA methyltransferase activity (Schaefer *et al.,* 2010).

### 1.3.1 DNMT1

DNMT1 is a 183kD protein responsible for the maintenance of methylation during replication and repair. *In vitro,* it has been shown to have a preference for hemi-methylated DNA over unmethylated substrates thus allowing it to copy the methylation pattern of the parental strand to the newly synthesized daughter strand (Jeltsch 2006). Even though it is referred to as the "maintenance" enzyme, there is increasing evidence that DNMT1 may also be involved in *de novo* methylation. For example, DNMT1 induced *de novo* methylation of CpG islands has been observed in human colon cancer cells (Jair *et al.,* 2006). DNMT1 is essential for maintaining epidermal progenitor cell function and its loss leads to premature differentiation (Sen *et al.,* 2010). Two fully functional isoforms of DNMT1 exist**.** The longer form is found in somatic and embryonic stem (ES) cells, and the shorter form in oocytes and pre-implantation embryos (Gaudet *et al.,* 1998). DNMT1 has been shown to be essential for cell survival (Egger *et al.,* 2006) and mutation of DNMT1 results in genome wide hypomethylation and embryonic lethality in mice (Li *et al.,* 1992). DNMT1 also seems to be

essential for the viability of cancer cells, for example, knockout of this enzyme results in HCT116 cell death (Chen *et al.,* 2007). In many cancers, for example, cervical cancer (Sawada *et al.,* 2007), hepatocellular carcinomas (Saito *et al.,* 2003) and stomach cancer (Etoh *et al.,* 2004), DNMT1 is over-expressed. However, down-regulation of DNMT1 may also contribute to carcinogenesis by causing hypomethylation. For example introduction of a hypomorphic allele of DNMT1 in mice has been shown to cause genomic hypomethylation along with the accelerated development of sarcomas (Eden *et al.,* 2003), and the carcinogenic agent cadmium, causes down-regulation of DNMT1 in a mouse testicular cell line (Singh *et al.,* 2009).

### 1.3.2 DNMT3 family

DNTM3A and DNMT3B are the active *de novo* methyltransferases responsible for cytosine methylation at previously unmethylated CpG sites. Both enzymes have a conserved PWWP domain responsible for their methyltransferase activity.

DNMT3A and DNMT3B can act independently of one another. For example, DNMT3B but not DNMT3A, has been shown to be responsible for the methylation of centromeric minor satellite repeats (Okano *et al.,* 1999). DNMT3A, but not DNMT3B, can restore methylation of the imprinted genes, XIST and H19 (Chen *et al.,* 2003). Although both enzymes are essential for male gametogenesis, they are uniquely responsible for methylation at different stages of this process (Turek-Plewa and Jagodzinski 2005).

On the other hand, there are circumstances in which the methylation of cellular genes is dependent on the contemporaneous binding of both DNMTA and DNTM3B to the same sequence (Chen *et al.,* 2003). For example, deletion of both DNMT3A and DNMT3B

catalytic domains is followed at certain loci by loss of methylation; this loss does not occur when the catalytic domain is deleted from only one of these enzymes (Okano *et al.,* 1999). Although we know both DNMT3A and DNMT3B are localized to the nucleus, it is still unclear whether their activity is dependent on specific DNA target sequences (Jones and Liang 2009). Mechanisms that trigger *de novo* methylation are also unknown. One possibility is that these enzymes are recruited by its PWWP domain to target genes; another, that their recruitment is dependent on binding to transcriptional factors such as MYC which has been shown to interact with DNMT3A leading to the repression of CDKN1A. It has also been suggested that RNAi systems may be involved in the targeting of the DNMTs but as yet this mechanism has not been fully defined (Klose and Bird 2006).

#### 1.3.2.1 DNMT3A

The DNMT3A gene is located on chromosome 2p23 and the protein is found in both the cytoplasm and nucleus where its expression is developmentally regulated. Alternative splicing of DNMT3A results in multiple transcript variants that encode four different isoforms. The DNA binding domain in DNMT3A is 50 residues long, the protein can dimerize, thus doubling the DNA binding surface and allowing it to methylate in one binding event, two CpGs separated by one helical turn (Klimasauskas *et al.,* 1994). DNMT3A is expressed at low levels in somatic cells, and at high levels in embryonal stem cells. Knockdown of DNMT3A inhibits melanoma growth and metastasis in mouse melanoma models suggesting that it may have an essential role in the pathogenesis of this tumour (Deng *et al.,* 2009).

#### 1.3.2.2 DNMT3B

The DNMT3B gene is located on chromosome 20q11.2. Six alternatively spliced transcript variants have been described although the full length sequences of variants 4 and 5 have not been determined. DNMT3B over-expression has been reported in breast cancer (Butcher and Rodenhiser 2007) and in colorectal cancer (Kanai *et al.,* 2001), and an increased risk of cancer at several sites has been associated with a polymorphism in this enzyme (Hu *et al.,* 2010).

Immunodeficiency, centromere instability and facial anomalies syndrome (ICF) is characterised by and is associated with point mutations in DNMT3B (Hansen *et al.,* 1999). DNA hypomethylation of pericentromeric heterochromatin is characteristic of the ICF syndrome. Commonly demethylated regions include satellite 2 and 3 repeats, cancer testes antigens and the inactive X chromosome (Jin *et al.,* 2008). A comparison of LCLs derived from healthy patients and those derived from patients with ICF syndrome has led to the identification of 800 genes which are potential DNMT3B targets. This study also suggested a possible interaction between the DNA and histone methylation systems. In ICF LCLs, loss of DNMT3B expression was associated with loss of H3K27me3, and reduced binding of a polycomb protein, SUZ12, to de-repressed genes (Jin *et al.,* 2008).

#### 1.3.2.3 DNMT3L

DNMT3L lacks those residues in the C-terminal domain which are required for methyltransferase activity and its contribution to DNA methylation is dependent on its

cooperation with the *de novo* DNMTs. DNMT3L has been shown to colocalize and co-immunoprecipitate with DNTM3A and DNMT3B (Hata *et al.,* 2002), and to enhance the activity of both enzymes (Kareta *et al.,* 2006; Suetake *et al.,* 2004). Disruption of DNMT3L in mice results in the failure to establish maternal methylation imprints (Bourc'his *et al.,* 2001).

## 1.4 DNA demethylation

Although it has yet to be established as to whether demethylation of DNA regulatory sequences occurs in a passive or active manner, various proteins have been suggested as possible DNA demethylases. One early candidate, MBD2 (Zhang *et al.,* 1999) was dismissed when Hendrich *et al.,* 2001 discovered that mice lacking MBD2 had normal patterns of DNA methylation and a normal phenotype. Although Barreto *et al.,* 2007 suggested that GADD45a might be a DNA demethylase, Jin *et al.,* 2008 could not confirm this. However, there is evidence for DNA demethylases in plants where it has been shown that lack of ROS1, DML2 and DML3 leads to the accumulation of methylated cytosines (Zheng *et al.,* 2008). More recently it has been suggested that DNMT3A and DNMT3B might initiate DNA demethylation (Metivier *et al.,* 2008; Kangaspeska *et al.,* 2008). Both studies suggest that DNMT3A and DNMT3B are involved in a dynamic cyclical recruitment of methylation to the promoters of oestrogen receptor target genes.

## 1.5 DNA methylation and regulation of transcription

DNA methylation, histone modifications and chromatin remodelling all contribute to the regulation of gene expression. This regulation is mediated by changes in chromatin state which can either be "open and active" or "closed and silent" depending on the interactions between transcription factors, cis-acting elements, and epigenetic modifications of the DNA sequence. In this thesis, I will focus on how changes in virus induced DNA methylation modulate changes in gene expression and therefore I now discuss in some detail the mechanisms which potentially regulate these changes.

### 1.5.1 Mechanisms of DNA methylation induced transcriptional change

In general, hypermethylation has been associated with gene repression and a closed chromatin structure, whereas hypomethylation has been associated with an open or loose chromatin structure resulting in gene activation. A number of mechanisms have been suggested to explain how the presence or absence of methyl groups can modulate gene expression (Bird *et al.,* 2002).

First, it has been suggested that the binding of transcription factors to gene promoters may prevent the addition of methyl groups, for example E2F, AP-2, c-myc, CREB/ATF and NF-kB have all been shown to bind unmethylated promoters but not methylated sequences. Consistent with these observations, Gebhard *et al.,* 2010 found that CpG islands that remain unmethylated in normal and in malignant cells contain specific sequence motifs that are identical to the consensus sequence for general TFs. These investigators went on to show that when these sites were stably bound by TFs in normal cells they are highly resistant to *de novo* methylation. Genes marked by methylation of histone H3 lysine 4 (H3K4) are usually transcriptionally active, and it has been suggested that this histone mark can also protect gene

promoters from *de novo* DNA methylation (Weber *et al.,* 2007; Appanah *et al.,* 2007). However, DNA methylation may not always act as a barrier, the viral transcription factor, BZLF1, preferentially binds methylated CpGs (Bhende *et al.,* 2004).

The transcriptional silencing of methylated genes has also been attributed to the activity of the methyl-binding proteins. Complexes which include either the methyl-CpG-binding protein 1 (MECP1) or MECP2 and one or more of the methyl binding domain proteins (MBD1, MBD2, MBD3, MBD4 and KAISO), have been shown to bind preferentially to methylated DNA and to inhibit transcription (Nan *et al.,* 1993). The binding of these complexes either limits access to transcription factors or results in the recruitment of transcriptional repressors or repressive complexes such as histone deacetylase complexes, chromatin remodelling complexes, or polycomb proteins (Wade 2001). More recently, the zinc finger and BTB domain containing proteins, ZBTB4 and ZBTB38, have been shown to repress transcription in a methylation dependent manner (Sasai and defossez 2009; Filion *et al.,* 2006).

A third model suggests that the impact of DNA methylation on gene expression may be mediated by changes to the nucleosome. For example, DNA methylation may result in a change in nucleosome positioning which may in turn result in a change in gene expression (Kass *et al.,* 1997). Alternatively, incorporation into the nucleosome of histone variants such as H3.3 and H2A.Z may prevent DNA methylation (Zilberman *et al.,* 2008).

## 1.6 Contribution of DNA methylation changes to carcinogenesis

Both increased DNA methylation and global reductions in DNA methylation are early epigenetic change contributing to carcinogenesis. I now briefly describe the evidence linking these changes to carcinogenesis before discussing how epigenetic therapy might be used in the treatment of malignancy.

### 1.6.1 Methylation changes in cancer

Although increased methylation at certain CpG islands has been observed in almost all cancers, global levels of DNA methylation have been shown to decrease with tumour progression (Feinberg and Vogelstein 1983; Laird 2005). Cancer cells have 20-60% less methylation than their normal counterparts (Esteller *et al.,* 2005) and this loss is consistently found in repetitive elements and in the introns of genes. Loss of methylation may contribute to tumourigenesis in several ways: loss of imprinting (Sakatani *et al.,* 2005); generation of chromosomal instability (Eden *et al.,* 2003; Daskalos *et al.,* 2009); reactivation of transposons (Bestor 2005); and reactivation of normally methylated oncogenes (Kim *et al.,* 2006; Honda *et al.,* 2004)

Hypermethylation of CpG islands and the associated silencing of tumour suppressor genes (TSG) have been found in both the early and late stages of tumourigenesis. The list of TSGs known to be methylated in cancer is now substantial but no single gene has been identified which is always methylated in cancer (Sharma *et al.,* 2010). Epigenetic silencing of tumour suppressor genes can lead to tumour initiation either by providing the second hit in the Knudson's "two-hit" model (Jones and Laird 1999) or through the silencing of both alleles. As to why some genes are hypermethylated in cancer and others not is unclear. However, it has been reported that regions which are hypermethylated in cancer are significantly more

likely to be marked by H3K27me3 in human embryonic stem cells (Ohm *et al.,* 2007; Schlesinger *et al.,* 2007; Widschwendter *et al.,* 2007) suggesting that the permanent silencing of these developmental genes may be important in tumourigenesis.

### 1.6.2 The use of epigenetic agents in the treatment of cancer

Unlike genetic aberrations, epigenetic changes are potentially reversible following therapeutic intervention. To date, the most commonly studied intervention has been the use of DNMT inhibitors. A number of these are now being tested in cancer. 5-azacytidine (azacitidine) and 5-aza-2'-deoxycytidine (decitabine) have been approved by the US food and Drug Administration (FDA) for the treatment of myelodysplastic syndrome. However, the toxicity associated with these drugs is not trivial. 1-(Beta-d-ribofuranosyl)-1,2-dihydropyrimidin-2-one (zebularine) may be a more effective anticancer drug as it appears to be more selective for cancer cells (Kantarjian *et al.,* 2003). All of these drugs act as a substrate for the DNMT enzymes and their use is associated with a global decrease in DNA methylation as a result of the passive demethylation of the genome which occurs after each round of cell division (Tao and Robertson 2003). However, when these different demethylating agents were used to "treat" an acute myeloid leukaemia cell line, they were found to have different effects on its transcriptional profile (Flotho *et al.,* 2009).

## 1.7 Arginine methylation

Histone proteins are susceptible to methylation, acetylation, ubiquitylation, sumoylation and phosphorylation on residues protruding from the N-terminal tail. The transcriptional consequences of these changes are dependent on the residue modified and the nature of the modification. For example, trimethylation of lysine 4 on histone H3 (H3K4me3) is associated with the activation of gene promoters (Liang *et al.,* 2004), whereas trimethylation of H3K9 and H3K27 is associated with transcriptional repression (Kouzarides 2007).

In this thesis I will focus on one particular histone modification, post-translational arginine methylation. Post-translational arginine methylation is important for regulating cellular processes such as signalling, gene transcription, mRNA splicing and DNA repair (Bedford and Clarke 2009). Nine protein arginine methyltransferases (PRMTs) have been identified in humans. PRMTs methylate and regulate not only transcription factors (TF) but also co-regulators and histones. Examples of TF regulated directly by PRMTs include p53, NF-κB, and RUNX1 (Zhao *et al.,* 2008). Co-regulators modulated by the PRMTs include p300 and CBP. PRMTs can also methylate histone tails and in so doing directly regulate gene expression.

Protein arginine methylation is also important in the replication of oncogenic viruses and for the transcriptional activation and stabilisation of viral proteins. Both PRMT1 and PRMT5 bind to the EBV protein EBNA1 which is important for the replication and mitotic segregation of viral genomes, and which is re-localised following inhibition of PRMT1 (Shire *et al.,* 2006). Arginine dimethylation of another EBV oncoprotein, EBNA2, is necessary for its efficient association with DNA bound transcription factors and with other viral promoters (Gross *et al.,* 2010). These examples illustrate how viruses can exploit the cell's arginine methylation machinery. However, with the exception of the Scophthalmus maximus

rhabdovirus which has been shown to up-regulate a fish PRMT1, virus-induced de-regulation of the PRMT has not been reported. I now describe in some detail three PRMTs (PRMT1, PRMT5 and CARM1) which are considered further in this thesis.

### 1.7.1 PRMT1

PRMT1, an 80 kDA protein, is the most abundant of the PRMTs and has been shown to be responsible for up to 85% of the total arginine methylation activity in RAT1 fibroblast cells and in mouse liver (Tang *et al.,* 2000). PRMT1 regulates gene transcription by methylating histone 4 and also Spt5, a transcriptional elongator which interacts with RNA polymerase II (Zhang and Chen 2003). The regulation of gene transcription by PRMT1 is sometimes dependant on its co-operation with other PRMTs (Kleinschmidt *et al*., 2008). For example, both CARM1 and PRMT1 are necessary for steroid hormone receptor-mediated reporter gene activation (Bedford and Richard 2005). Whereas, knockdown of both CARM1 or PRMT1 has been shown to result in the down-regulation of STAT5-controlled genes, knockdown of only one of these genes had no effect (Kleinshmidt *et al.,* 2008).

### 1.7.2 CARM1

CARM1, also known as PRMT4, has restricted substrate specificity (Lee and Bedford 2002). CARM1 recruitment to transcriptional promoters results in elevated levels of methylation on the $17^{th}$ and $26^{th}$ arginine residues of histone 3 (H3R17 and H3R26, respectively); both of these methylation marks are associated with transcriptional activation. In addition to the methylation of histones, CARM1 also methylates other transcriptional co-activators for example, SNRPB, a small nuclear protein involved in pre- mRNA splicing (Cheng *et al.,* 2007).

### 1.7.3 PRMT5

PRMT5 functions in several nuclear and cytoplasmic complexes. It plays an important role in the control and modulation of gene transcription, regulating for example, the expression of IL2 and CCNE1 (Richard *et al.,* 2005; Fabbrizio *et al.,* 2002). Homozygous PRMT5 deletions in mice lead to death of zygotes (Krause *et al.,* 2007). PRMT5 can act as a transcriptional co-activator or co-repressor. For example, it acts as a transcriptional co repressor when bound to COPR5, a histone binding protein; this complex preferentially methylates specific arginine residues on histone 4 (Lacroix *et al.,* 2008). Alternatively, PRMT5 can act as a co-activator when it binds to the SWI/SNF family (Pal *et al.,* 2003, Dacwag *et al.,* 2007). PRMT5 has also been shown to repress PRDM1, the master regulator of B cell differentiation (Ancelin *et al.,* 2006).

### 1.7.4 Links between PRMT and DNA methylation

Recently a link between the PRMTs and DNA methylation has been established. It has been demonstrated that the symmetric di-methylation of histone H4 arginine 3 (H4R3me2s) by PRMT5 is required for the subsequent DNA methylation of the human β-globulin locus (Zhao *et al.,* 2009). The investigators concluded that DNMT3A was responsible for this methylation event as knockdown of PRMT5 led to a reduction in DNMT3A binding, loss of DNA methylation and gene activation.

## 1.8 Epstein-Barr virus

EBV, the first DNA tumour virus to be discovered, is a member of the family *Herpesviridae*. Although more than 130 herpesviruses have been identified, only eight have been isolated from humans and are divided into three subfamilies, the *Alphaherpesvirinae*, the *Betaherpesvirinae* and the *Gammaherpesvirinae* (Table 1.1).

| Herpes sub-family | Type | Disease |
|---|---|---|
| **Alpha** | Herpes simplex virus-1 | Recurrent facial herpetic lesions |
| | Herpes simplex virus-2 | Recurrent genital herpetic lesions |
| | Varicella-zoster virus | Chicken pox (primary)<br>Shingles (reactivation) |
| **Beta** | Human cytomegalovirus | Typically asymptomatic but can be a complication in immunosuppressed patients |
| | Human herpes virus-6 | Mild childhood roseola |
| | Human herpes virus-7 | Mild childhood roseola |
| **Gamma** | Epstein-Barr virus | Asymptomatic<br>Infectious mononucleosis<br>Associated with a subset of cancers |
| | Kaposi sarcoma associated herpesvirus | Associated with a subset of cancers |

Table 1.1: The human herpesviruses

All herpesviruses share four biological properties:

- they use a large number of enzymes for protein processing and DNA synthesis
- the production of progeny virus destroys the infected cell
- synthesis of viral DNA occurs in the nucleus
- each is able to establish a latent life-time infection in their natural host.

Although they share these biological properties they differ in terms of their genomes. In particular, their nucleotide composition varies considerably, for example, HHV-6 has a G+C content of 43%, HSV-1 of 68% and EBV of 60% (Weir 1998). They also differ in terms of CpG methylation which is believed to be related to the location of virus replication (Weir 1998). The alphaherpesviruses, which establish latency in non-dividing neuronal cells lack CpG methylation, they do not undergo replication and methylation as a result of cell division (Dressler *et al.,* 1987). The betaherpesviruses exhibit a local CpG methylation over the major immediate early region. The specific CpG suppression observed in the major immediate-early region of the betaherpesviruses is not yet understood, but may be the result of localized exposure of this region to methylases during latency (Gompels *et al.,* 1995). Finally, EBV and other gammaherpesviruses that are maintained as episomes in proliferating lymphoblastoid cells may be routinely methylated by host cell enzymes during multiple cell divisions, which may aid in allowing the virus to go undetected by the immune system (Tao and Robertson 2003, Gunther and Grundhoff 2010).

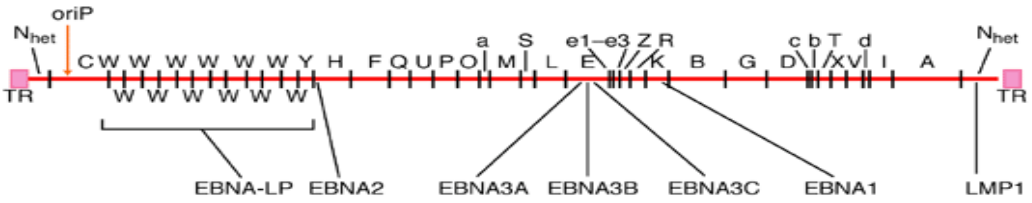
Only two members within the gamma herpesvirus subfamily infect humans, EBV and Kaposi's sarcoma-associated herpesvirus (KSHV), both have been linked to human

malignancy. Within the gamma herpesviruses, EBV belongs to the Lymphocryptovirus (LCV) genus, members of which display lymphoid cell tropism and are oncogenic. An EBV-related virus has also been isolated from the spontaneous B cell lymphomas arising in common marmosets. This virus termed marmoset LCV or Callitrichine herpesvirus 3 was first identified in animals with spontaneous B cell lymphomas at the Wisconsin National Primate Research Center. Subsequent work has revealed that 40-60% of captive marmosets are seropositive for this virus with most animals not revealing overt signs of clinical disease. Definition of the viral gene repertoire revealed collinear genomic organization and 60 open reading frames with homology to those seen in EBV and other primate LCVs (Rivailler *et al.,* 2002). However, whether this virus is subject to methylation similar to EBV is still unknown.

### 1.8.1 Virus structure

EBV has a toroid-shaped protein core which is wrapped in DNA and surrounded by a nucleocapsid consisting of 162 capsomeres. The nucleocapsid is found inside a protein tegument located between an inner and an outer envelope from which project external glycoproteins (Kieff and Rickinson 2001). The EBV genome is composed of linear double stranded DNA, approximately 184 kilobase pairs (kbp) in length. As the EBV genome was sequenced from a *Bam*HI fragment library, its genes, their promoters, open reading frames, and polyadenylation sites are defined by their position on the *Bam*HI restriction fragment map and named in a descending order of size, with A being the largest (Figure 1.3).

A




B

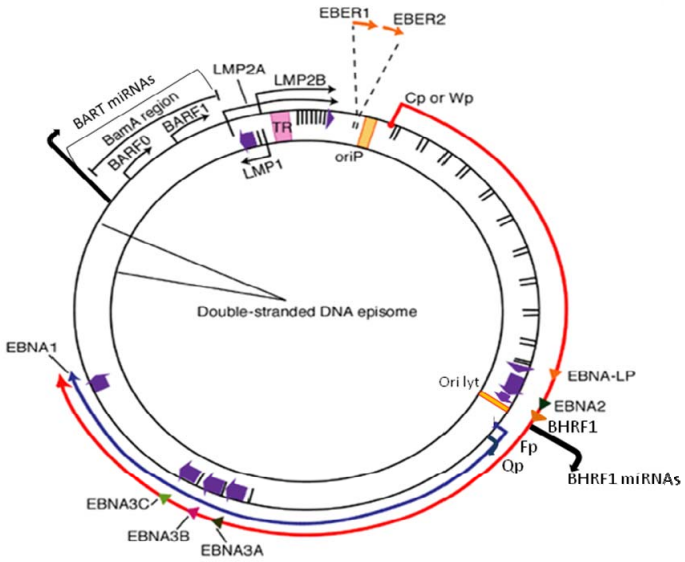



**Figure 1.3 The EBV genome. A.** Diagram showing the location of open reading frames for the EBV latent proteins on a *Bam*HI restriction endonuclease map of the prototype B95.8 EBV genome. **B.** Diagram showing the location and transcription of the EBV latent genes on the double-stranded viral DNA episome. The origin of plasmid replication (oriP) and lytic replication (ori lyt) are shown in orange. The large solid blocks (in purple) represent coding exons for each of the latent proteins and the arrows indicate the direction in which they are transcribed (modified from Young and Murray 2003).

### 1.8.2 Repeat regions of the EBV genome

The EBV genome is characterised by tandem reiterated 0.5 kbp repeat sequences which are called terminal repeats (TRs), and which are located at both ends of a linear genome (Figure 1.3). Fusion of these TRs following EBV infection of B cells results in the circularisation of viral DNA and the formation of a closed episome (Kintner and Sugden 1979). Quantitation of TRs can be used to determine the clonality of EBV infected cell populations because different infectious events will give rise to latent episomes with varying numbers of TRs which remain constant in each subsequent generation (Given *et al.,* 1979). Also within the EBV genome is a large tandem repeat region known as the major internal repeat 1, IR1; this region comprises numerous repeats of a 3072bp BamHI W region containing one of the EBNA promoters, Wp. The number of tandem repeats within IR1 varies among EBV strains and their progeny (Allan and Rowe 1989); for example, the prototype B95.8 EBV genome has 11 of these repeats as shown in the linear map in Figure 1.3a.

### 1.8.3 EBV strains

EBV isolates can be classified into two distinct families, types 1 and 2 (these were once referred to as type A (B95.8-like) and type B (AG876-like). The difference between the two types is based on variations in the sequence of their EBV nuclear antigens, EBNA2, EBNA3 and EBNA-leader protein (EBNA-LP) (Sample *et al.,* 1990). The most common EBV strain is type1, with type 2 found mainly in equatorial Africa or in people infected with human immunodeficiency virus (HIV) (Falk *et al.,* 1997).

### 1.8.4 Virus binding, entry and circularisation

*In vitro* studies have shown that early in the course of primary infection, EBV infects B cells by binding the CD21 receptor which is expressed almost exclusively in B lymphocytes. Two interactions, one between the viral glycoprotein gp 350/220 and CD21, and another, between the viral envelope protein complex gp42/gH/gL and the major histocompatability complex (MHC) class II molecules, allow EBV to attach to and to penetrate B cells (Molesworth *et al.,* 2000). Following endocytosis of the virus into the host cell, the genome circularises to produce the covalently closed extrachromosomal episome characteristic of latently infected cells (Figure 1.5B). Episomes are maintained by the host cell DNA polymerase which replicates EBV episomes once per cycle in early S phase (Adams 1987). Amplification of the viral genome occurs early after transformation and latently infected cells harbour numerous viral episomes in their nucleus (Kinter and Sugden 1979).

## 1.9 Primary EBV infection

EBV infects approximately 95% of the world's adult population. The virus has at least two potential target cells, B-cells and oropharyngeal epithelial cells. As EBV is transmitted orally, the oropharynx is normally the primary site of infection but whether the primary target is a B cell or an epithelial cell is unknown. Children who become infected with EBV are generally asymptomatic or show vague symptoms indistinguishable from those associated with other childhood illnesses. If infection occurs in adolescence or adulthood it may cause infectious mononucleosis (IM) which is characterised by sore throat, fever and fatigue. In IM, virus induced activation of B cells results in the enlargement and ulceration of the tonsils. Following primary infection, EBV persists in the infected host as a lifelong asymptomatic infection. To achieve life-long persistence, EBV colonises the B cell pool where it establishes a latent infection which is characterised by the expression of a limited subset of virus genes known as "latent" genes (Thorley-Lawson 2001). I next describe each of these viral latent genes before commenting on the latent states with which they are associated.

## 1.10 Latency

In its different latent states, EBV produces a limited number of "latent proteins" which include the EBV nuclear antigens EBNA1, 2, 3A, 3B, 3C and LP; the latent membrane proteins (LMP)1, 2A and 2B; the *Bam*HI A RNA, BARF1; the EBV-encoded RNAs, EBER 1 and EBER 2; and BART RNA transcripts. EBV also codes for at least 23 miRNAs that are expressed in latently infected cells (Yun Zhu *et al.,* 2009). The contribution of each of the latent genes to viral maintenance and transformation is shown in Table 1.1.

### 1.10.1 Latency states during viral persistence *in vivo*

EBV persists within the B cell compartment in four different latent states dependent on the stage of B cell development. In each of these four latency states, "Latency 0", "Latency I", "Latency II" and "Latency III" a different pattern of viral gene expression is observed. Initial infection of B cells with EBV activates the Latency III transformation program during which all of the EBV-latent genes are expressed. This pattern of gene expression is found in most cases of Post Transplant Lymphoproliferative disease (PTLD). Following the activation of the growth transcription program, the cells migrate to the germinal centre (GC) where they continue to proliferate and mature but fail to expand in numbers (Roughan *et al.,* 2010). Here, these cells express the Latency II genes, the EBERs, BARTs, EBNA1, LMP1 and LMP2; this pattern is also seen in Hodgkin's lymphoma (HL) and in nasopharyngeal carcinoma (NPC) (Young *et al.*, 1988). In response to survival signals, these cells exit the GC as resting memory B cells where they do not express any EBV proteins (Latency 0). When resting memory cells divide the Latency I genes, the EBERs, BARTs and EBNA are expressed, a pattern also seen in Burkitt's lymphoma (BL) (Rowe *et al.*, 1987). These cells eventually return to the tonsil where they may undergo plasma cell differentiation which can

trigger viral replication. The virus is then either shed from the host through saliva, or it can re-infect other B cells, or both (Figure 1.4).

| | |
|---|---|
| **EBNA1** | EBNA1 is required for the replication and maintenance of the episomal genome (Lee *et al.,* 1999). |
| **EBNA2** | EBNA2 plays a key role in B-cell growth transformation by initiating and maintaining the proliferation of infected B-cells upon EBV infection *in vitro*. (Skare *et al.*, 1985; Farrell *et al.,* 2004). |
| **EBNA3 family** | *In vitro* studies with EBV recombinants have demonstrated that EBNA3A and EBNA3C (Tomkinson *et al.*, 1993) are essential for B-cell transformation, whereas EBNA3B is not (Tomkinson and Kieff, 1992). |
| **EBNA-LP** | This protein is not essential for B cell transformation *in vitro* but is required for the outgrowth of LCLs (Allan *et al.,* 1992). |
| **LMP1** | LMP1 behaves as a classical oncogene and is essential for B cell transformation *in vitro* (Kaye *et al.*, 1993). It is also responsible for the |

| | |
|---|---|
| | activation of at least four signalling pathways; the nuclear factor-kappa B (NF-κB) pathway (Huen *et al.*, 1995); the c-Jun N-terminal kinase (JNK)/activator protein (AP)-1 pathway (Kieser *et al.*, 1997); the p38/mitogen-activated protein kinase (MAPK) pathway (Roberts and Cooper, 1998); and the JAK/STAT pathway (Gires *et al.*, 1999). |
| **LMP2** | The gene encoding LMP2 yields two distinct proteins LMP2A and LMP2B (Longnecker and Miller, 1996). Neither LMP2A nor LMP2B are essential for B-cell transformation *in vitro* (Longnecker, 2000). LMP2A is capable of driving proliferation and survival of B cells in the absence of signalling through the BCR (Caldwell *et al.*, 1998). |
| **BARF1** | BARF1 acts as an oncogene (Hayes *et al.*, 1999) but is dispensable for B-cell transformation (Cohen and Lekstrom 1999). |
| **EBERs** | The EBV-encoded RNAs, EBER 1 and 2 are expressed in all latency states. They are not essential for EBV-induced transformation but can enhance immortalisation of B lymphocytes (Yajima *et al.,* 2005). |
| **BARTs** | The BARTs are a group of RNAs that may be involved in the maintenance of the latent state (van Beek *et al.*, 2003). |
| **EBV-encoded miRNAs** | EBV encodes 23 miRNAs identified to date (Grundhoff *et al.*, 2006). These are arranged in two clusters in the EBV genome, one located in the introns of the BART gene, and the other adjacent to the *Bam*HI fragment H rightward open reading frame 1 (BHRF1) gene. These miRNAs are thought to play a role in the regulation of gene expression (Cai *et al.*, 2006; Pfeffer *et al.*, 2004). |

**Table 1.2: Latent genes:** The contribution of each of the latent genes to viral maintenance and transformation is described.

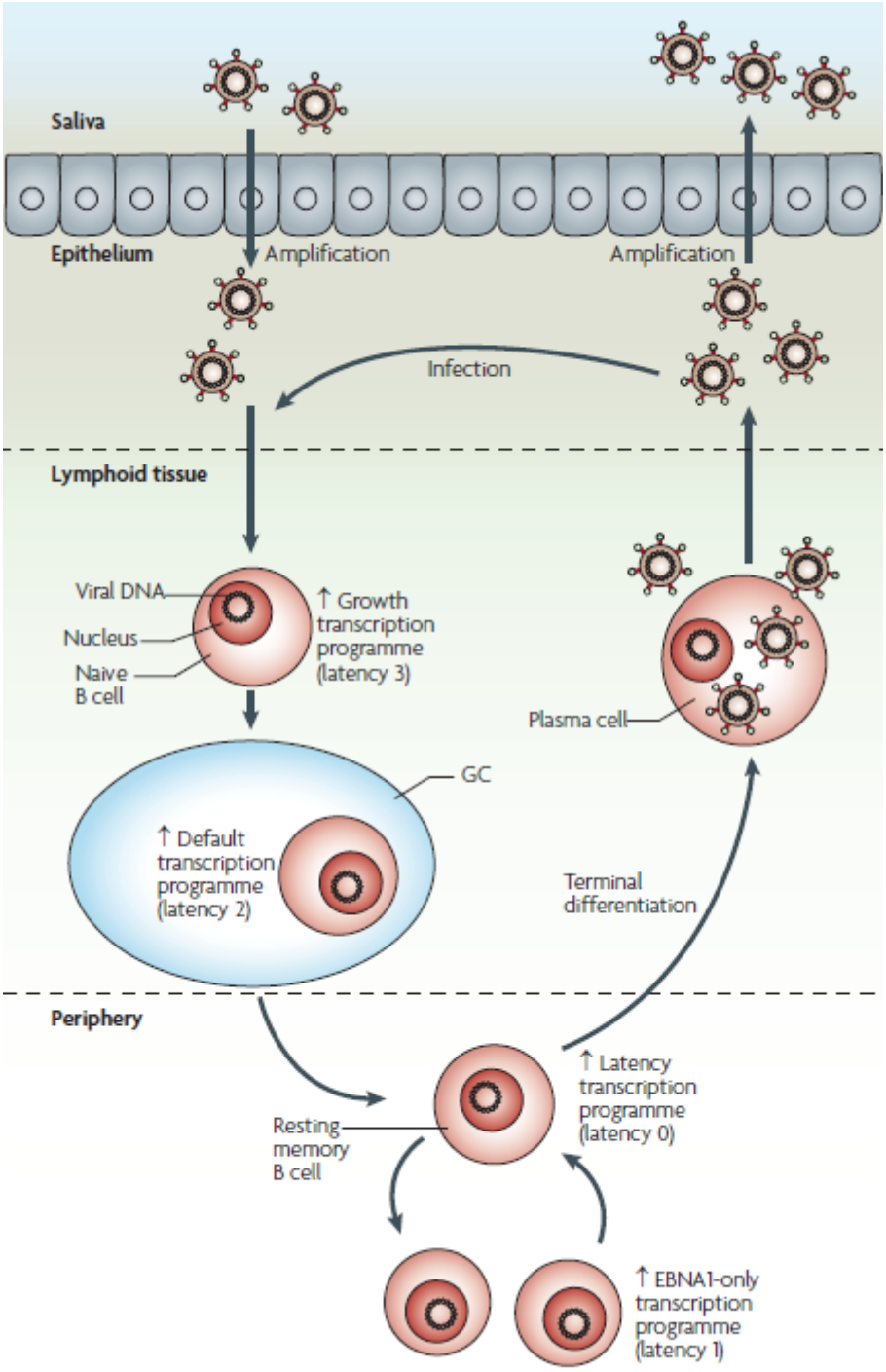



**Figure 1.4 The current model of viral infection and persistence:** EBV enters the epithelium where it probably initiates a lytic infection. It may then spread to B-cells where it initiates a latency III transformation program. EBV-infected B-cell numbers are controlled by a latent-antigen-specific primary-T-cell response but a fraction of these cells down-regulate antigen expression to evade immune recognition. Activation of the growth transcription

programme allows the cells to migrate to the GC where they activate a “default transcription programme” (Latency II). Rescue or survival signals allow the cell to exit the GC as a resting memory cell and the “latency transcription programme” is initiated (Latency 0). When these cells divide they express the Latency I programme of gene expression. The cells eventually return to the tonsil where they may undergo plasma cell differentiation which can trigger viral replication. (Diagram from Thorley-Lawson and Allday 2008).

## 1.11 Viral gene expression following primary infection *in vitro*

EBV is the most effective transforming agent known for human cells *in vitro*, and infection of resting human B lymphocytes in culture consistently produces permanently proliferating lymphoblastoid cell lines (LCLs) (Henle *et al.,* 1967). These lines express the full range of latency III viral proteins causing B cells to enter the cell cycle and proliferate. Thereafter they undergo a change in their cellular phenotype and begin clumping together with the expression of adhesion molecules and activation markers (Alfieri *et al.,* 1991). In terms of their growth and expression of activation, adhesion and differentiation markers, these EBV transformed LCLs resemble uninfected B cells proliferating in response to antigen and T-cell help (Hollyoake *et al.,* 1995). Although the majority of cells are in latency state III, a small proportion of cells enter the lytic cycle, producing virus particles (Kieff and Rickinson 2007).

### 1.11.1 Promoter regulation of viral gene expression following primary infection *in vitro*

The first viral promoter activated following infection of primary B cells with EBV is the *Bam*HI W promoter, Wp, from which the transcripts for all six EBNAs are initiated. Although it is the first promoter to drive EBNA expression, this function is soon taken over by Cp, the dominant promoter in most LCLs (Schlager *et al.,* 1996). Approximately 24 hours after *in vitro* infection of B-cells with EBV, the first viral genes are expressed - EBNA-LP and EBNA2 (Alfieri *et al.*, 1991). Subsequently, EBNA2 transactivates LMP1, LMP2 and other cellular genes including CD21 and CD23, a B-lymphocyte activation antigen (Wang *et al.*, 1990). By 32 hours, all the EBNAs, and LMP1 are expressed and CD21 and CD23 are further up-regulated. The mechanism for the onset of LMP2A and LMP2B expression still remains unknown (Longnecker and Kieff, 1990). In Latency I and II, EBNA1 is transcribed from the

*Bam*HI Q (Qp) promoter rather than Wp or Cp. The alternate use of promoters during different latency stages may provide the virus with a greater flexibility in controlling transcription. Understanding the virus' strategy for persistence *in vivo* may help us understand more about EBV-associated diseases.

### 1.11.2 DNA methylation of promoters regulating viral gene expression *in vitro* and *in vivo*

Oncoviruses have developed different strategies for evading the host immune response, all of which are essentially aimed at camouflaging the virus in the host cell. By restricting the expression of certain viral genes and those associated with the immune response the virus can go undetected. One mechanism by which this might be achieved is by DNA methylation. (Tao and Robertson 2003; Niller *et al.,* 2008; Fernandez *et al.,* 2009).

DNA methylation and its involvement in the regulation of EBV latent promoters was first observed when latently EBV-infected human lymphoid lines were demethylated using 5-aza-cytidine (Ben-Sasson and Klein 1981, Szyf *et al.,* 1985). This was followed by various methylation studies of EBV-positive BL cell lines (Allday *et al.,* 1990, Ernberg *et al.,* 1989, Jansson *et al.,* 1992) and since then the methylation profile of specific viral genes has been extensively described (Fejer *et al.,* 2008, Park *et al.,* 2007; Paulson and Speck 1999; Ambinder *et al.,* 1999; Takacs *et al.,* 2010).

Recently, a comprehensive methylation analysis of 94 transcriptional start sites of EBV was carried out on the EBV DNA methylomes from 22 different lymphoid samples; these included Hodgkin's lymphoma, non-Hodgkin's lymphoma, Burkitt's lymphoma primary tissue and cell lines, as well as LCLs (Fernandez *et al.,* 2009). This analysis revealed that only five promoters are consistently unmethylated in the genome; these include the Qp promoter and

those regulating EBER1 and EBER2, BZLF1 and LMP2B/LMP1. This study concluded that the EBV genome is extensively methylated in the majority of lymphomas and primary tumours but not in cells from IM patients and from LCLs. This study also revealed a functional association between viral methylation and transcription; the presence of hypermethylated EBV transcription start sites in cancer cells was associated with transcriptional silencing of neighbouring viral genes.

#### 1.11.2.1 Regulation of Wp

The first evidence that DNA methylation was involved in the regulation of Wp was found when treatment of the BL line, Rael-BL with 5-aza-deoxycytidine resulted in the expression of all six EBNAs (Masucci *et al.,* 1989). Since then, Wp has been shown to be methylated in BL cell lines displaying Latency I form of infection (Tierney *et al.,* 2000); the majority of EBV positive lymphomas (Fernandez *et al.,* 2009); PTLD tumours; and in the B cells of acute IM patients (Tierney *et al.,* 2000, Tao and Robertson 2003).

Wp remains unmethylated in freshly established LCLs. However, four weeks post infection Wp is methylated even though not all copies are silenced (Hutchings *et al*. 2006). It is has been suggested that DNA methylation may interfere with Wp expression by blocking the binding of BSAP, a transcription factor that has been shown to contribute to Wp activity, (Tierney *et al.,* 2000).

#### 1.11.2.2 Cp, Fp and Qp methylation

Cp and Fp, (a promoter that mediates the expression of EBNA1) are unmethylated and active in LCLs, in PTLD and in B cells of acute IM patients (Tierney *et al.,* 2000; Fernandez *et al.,*

2009), but hypermethylated and silent in NPC, HL, and BL (Fernandez *et al.,* 2009; Robertson *et al.,* 1996). Qp has never been found to be methylated regardless of its transcriptional activity and is thought to be regulated by another epigenetic mechanism (Tao *et al.,* 1998; Fernandez *et al.,* 2009).

#### 1.11.2.3 LMP promoter methylation

A number of observations suggest that LMP1 expression is regulated by methylation of its LRS promoter. In Latency I BLs and in NPC tumours where LMP1 is silent, the promoter is highly methylated. However in LMP1 expressing latency III cells and in all other lymphoma and primary tumours, the promoter is predominantly unmethylated (Minarovits *et al.,* 1994, Falk *et al.,* 1998, Salamon *et al.,* 2001, Takacs *et al.,* 2001, Fernandez *et al.,* 2009). The LMP2A promoter is methylated and inactive in the NPC cell line C-666.1 but remains unmethylated in LCLs and in the majority of lymphomas (Fernandez *et al.,* 2009).

## 1.12 Lytic EBV infection

Completion of the EBV life-cycle is marked by the production of infectious virus. For this to happen, EBV must be reactivated from its latent state in B cells. Lytic infection is initiated by the expression of BZLF1 and BRLF1, viral transactivator proteins which trigger the initiation of a generative replication cycle in which cellular RNA, DNA and protein synthesis are repressed and virus DNA synthesis takes over. At the final stages of lytic cycle, icosahedral nucleocapsids are assembled, viruses are enveloped by budding out from the nuclear membrane, and the cell is lysed to release infectious enveloped virions containing linear genomes (Kieff and Rickinson 2007).

Lytic viral genes are expressed in three phases. The late (L) lytic genes are the last to be expressed 48-72 hours post-induction of the lytic cycle. EBV L lytic genes mostly code for structural proteins which can be grouped into glycoproteins and non-glycoproteins (Farina *et al.*, 2005). The glycoprotein group includes BILF2, BDLF3, BALF4, BXLF2 and BLLF1; and the non-glycoprotein group includes BCLF1, BNRF1 and BXRF1. The latter code for proteins of the nucleocapsid and tegument which form the major external virus component (Kieff and Rickinson 2007).

The transcription of the early (E) lytic genes of which there are at least 30 mRNAs, is independent of cellular protein and viral DNA synthesis (Biggin *et al.*, 1987). In early lytic infection, two abundant early nuclear proteins BMRF1 and BMLF1 are expressed (Wong and Levine, 1986), along with the BALF2 and BHRF1 early EBV proteins (Pearson *et al.*, 1987).

The immediate-early (IE) genes are the first lytic genes to be expressed and are now discussed in more detail because of their known regulation by methylation.

### 1.12.1 IE lytic genes

BZLF1 (Zta) and BRLF1 (Rta) are the IE lytic genes involved in initiating the switch from the latent state. Ectopic expression of BZLF1 alone has been shown to be sufficient to induce lytic cycle *in vitro* (Rooney *et al.*, 1989 Kudoh *et al.*, 2003). BZLF1 is a sequence specific DNA-binding protein which regulates the expression of viral and cellular genes by binding to Zta response elements (ZREs) found in their promoters (Heather *et al.,* 2009; Flower *et al.,* 2010). Whether BZLF1 binds to a ZRE or not depends on the methylation status of that ZRE; it will bind to some ZRE when they are unmethylated but only bind to others when they are methylated (Dickerson *et al.,* 2009; Bhende *et al.,* 2005; Karlsson *et al.,* 2008). For example, BZLF1 binds to and activates the methylated form of the BRLF1 promoter more efficiently than it does the unmethylated form (Dickerson *et al.,* 2009)

Although BZLF1 is expressed very early following infection in the majority of primary B cells, it fails to induce the EBV lytic phase at this time. It now appears that for BZLF1 to activate the gene expression cascade leading to lytic infection, it must first bind to a number of viral promoters involved in that cascade. However, BZLF1 can only bind to these promoters which are likely to be downstream of BRLF1 after they have become methylated (Kalla *et al.,* 2010) This inevitably delays the onset of lytic infection because the viral genome is unmethylated when cells are first infected, and only becomes methylated over time, presumably following virus-induced activation of the cell's DNA methylation machinery. To regulate its own life-cycle, EBV appears to have evolved a time-dependent epigenetic switch, one which can be tripped by host cell induced changes in the methylation status of viral promoters bound by BZFL1. However which of the host DNMTs binds to the BZLF1 promoter has yet to be determined and is one of the objectives of this thesis.

## 1.13 The impact of oncogenic viruses on the regulation of DNMTs and methylation of the cellular genome

In this section, I first introduce those viruses which have been implicated in the aetiology and pathogenesis of human malignancy. I next briefly review the evidence implicating HBV, KSHV, HTLV-1 and HPV in the de-regulation of the DNMTs, and the impact of this de-regulation on the viral and cellular epigenome. I then provide a brief description of the EBV associated malignancies and the contribution of EBV-associated methylation changes in cellular genes to the pathogenesis of these tumours. Finally, I review the prospects for using methylation inhibitors in the management of these tumours.

### 1.13.1 Introduction to oncogenic viruses

In recent years an increasing number of human cancers have been shown to be virus associated and it has been suggested that viruses have a causal role in 15% of all cancers worldwide (Pisani *et al.,* 1997). Five virus families are associated with human cancers; Herpesviruses, Hepadnaviruses, Papillomaviruses, Flaviviruses and Retroviruses each of which is ubiquitous and capable of establishing persistent infection (Flanagan 2007). Not only may the establishment of persistent infections be dependent on epigenetic changes in the viral genome, oncogenic viruses can also cause aberrant DNA methylation and other epigenetic changes in cellular genes (Flanagan 2007, Ferrari *et al.,* 2009).

#### 1.13.1.1 Hepatitis B virus

HBV is a DNA virus associated with hepatocellular carcinoma. Although methylation of the HBV genome increases from early to late stages of carcinogenesis (Fernandez *et al.,* 2009), that of the X gene, coding for the HBx protein, appears to remain unmethylated throughout the viral life cycle. HBX is responsible for the up-regulation of DNMT1 and DNMT3A and has been shown to cause the hypermethylation of the TSG, IGFBP3 (Park *et al.,* 2007). It has also been shown to interact with DNMT3A and to recruit it to the regulatory promoters of MT1F and IL4R which are then silenced.

#### 1.13.1.2 Kaposi's sarcoma-associated herpesvirus

KSHV, like EBV, is a persistent herpesvirus. DNA methylation regulates the viral gene promoter ORF50, thus ensuring its repression during latency (Shamay *et al.,* 2006). The main transforming KSHV gene, latency-associated nuclear protein (LANA), is unmethylated throughout the viral life cycle (Chen *et al.,* 2001) and has been shown to interact with a number of epigenetic regulators including the DNA methyl binding protein MeCP2, the mSin3 transcriptional repression complex and the histone methyltransferase SUV39H1, (Flanagan 2007). Like HBx, LANA can bind to and retarget DNMT3A. It preferentially relocalizes DNMT3A from the nuclear matrix into the chromatin fraction (Shamay *et al.,* 2006) and recruitment of *de novo* DNMTs by the virus has been shown to repress CDH13 and TGF-β cellular promoters by DNA methylation (Shamay *et al.,* 2006; Di Bartolo *et al.,* 2008).

#### 1.13.1.3 Human T-cell leukemia virus type 1

HTLV-1 is a single stranded RNA retrovirus associated with adult T-cell leukemia (ATL). Tax, the main transforming protein of HTLV-1, controls viral gene expression through long terminal repeats (LTRs). Several studies have found DNA hypermethylation of the 5'LTR is more frequent in ATL cells than in normal carriers (Taniguchi *et al.,* 2005). However, it is unknown which of the DNMTs is responsible for these changes.

#### 1.13.1.4 Human papilloma virus (HPV)

HPV is responsible for a variety of benign proliferations but infection with the high-risk mucosal HPV types, such as HPV16 and HPV18 cause squamous intraepithelial lesions of the cervix which may progress to invasive squamous cell carcinoma. Whereas both viruses are almost completely unmethylated in preimmortal keratinocytes, the viral genome in their immortal counterparts is densely methylated (Fernandez *et al.,* 2009). The virally encoded oncoprotein, E7 can associate directly *in vitro* and *in vivo* with DNMT1 through the E7 CR3 zinc finger domain, leading to increased activity of the enzyme (Burgers *et al.,* 2006). It has also been shown that HPV 16 and 18 infection up-regulates DNMT3B, and that this is associated with p16INK4a promoter hypermethylation. However, this association was only found in non-smoking lung cancer patients, and only then in females (Lin *et al.,* 2005); the viral gene responsible for this remains to be elucidated.

## 1.14 EBV-associated cancers

In this section, I provide a brief summary of the EBV-associated cancers focusing in the main on HL. In EBV-associated malignancies, the virus is almost exclusively found in a latent state. The pattern of virus latent gene expression depends upon the type of tumour and state of differentiation of the infected cell indicating that the contribution of EBV to carcinogenesis varies according to tumour type (Landais *et al.*, 2005). Diseases such as PTLD and endemic BL, are nearly 100% EBV-positive, suggesting an essential role for EBV. Other neoplasms such as HL and gastric carcinoma are less frequently associated with EBV. In these tumours, the contribution of EBV to malignant transformation may be replaced by other mutations in virus-negative disease.

### 1.14.1 Nasopharyngeal carcinoma

Nasopharyngeal carcinoma (NPC) is a tumour of the surface epithelium of the nasopharynx. NPC has a high incidence in China and many parts of Southeast Asia, intermediate in North and East Africa and a low incidence in the rest of the world (Epstein and Achong, 1979). There are three forms of NPC: keratinizing squamous cell carcinoma (WHO1) and non-keratinizing squamous cell carcinoma (WHO2/3) which is further divided into differentiated (WHO2) and undifferentiated (WHO3) ( Shah and Young 2009). Both WHO2 and WHO3 are associated with EBV; these tumours contain monoclonal viral genomes indicating that viral infection occurs before malignant expansion (Raab-Traub and Flynn 1986). It has been suggested that genetic alterations in pre-malignant cells facilitate stable maintenance of latent EBV infection which when followed by other transforming events, can drive that clonal expansion of the cells necessary for tumour formation (Young and Rickenson 2004). Thus, in NPC, EBV may act as a tumour promoting agent rather than an initiating factor.

EBV infection of human epithelial cells has been reported to be followed by the up-regulation of DNMT1, DNMT3B and DNMT3A. The EBV oncogene, LMP1, up-regulates DNMT1 in both un-transformed (MDCK) and transformed cells (NPC079) and has been shown to induce hypermethylation of the CDH1 tumour suppressor gene in both (Tsai *et al.,* 2002). The same group have also reported that the up-regulation of DNMT1 by LMP1 occurs through the c-Jun NH2-terminal kinase (JNK)-activator protein AP-1 signalling pathway (Tsai *et al.,* 2006). In NPC cells, LMP1 induces promoter hypermethylation of RAR-beta2 via the up-regulation of DNMT1, DNMT3A, and DNMT3B, leading to a decrease in RAR-beta2 expression (Seo *et al.,* 2008).

### 1.14.2 Gastric carcinoma

Approximately 10% of gastric cancers test positive for EBV. However, whereas the more common diffuse and intestinal types are rarely found to be EBV positive (< 10%), the less common lymphoepithelioma-type is found to be EBV positive in 90% of cases. EBV positive tumours are more common in women, in tumours of the gastric cardia, and in the stump remnant left after gastric surgery (Murphy 2009). Monoclonal EBV is present in every malignant cell in virus positive disease implying that infection occurred prior to clonal expansion (Imai *et al.,* 1994). In gastric carcinoma cell lines, LMP2A was found to activate DNMT1 through the phosphorylation of STAT3, resulting in loss of PTEN expression following methylation of its promoter (Hino *et al.,* 2009). A number of studies have reported substantial and significant differences in the frequency in which TSG are differentially methylated in EBV-positive and EBV negative tumours. TSG methylation is significantly more common when EBV is present.

### 1.14.3 Burkitt's lymphoma

Burkitt's lymphoma (BL) is a lymphatic tumour that occurs in both endemic and sporadic forms. The endemic form is associated with *Plasmodium falciparium* malaria (Morrow 1985) and is restricted to equatorial Africa; EBV is present in 95% of cases. The sporadic form occurs 10 to 100 times less frequently and EBV is present in only 15% of cases (Magrath 1990). BL is also common among adult HIV carriers in the developed world and often arises as the first AIDS-associated illness in relatively immunocompetent patients; 30-40% of these tumours are EBV-positive (Kieff and Rickinson, 2007). Irrespective of their EBV status, the phenotype of BL cells ($CD10^+$, $CD77^+$, $BCL6^+$) is remarkably similar to that of GC centroblasts, and coupled with detection of ongoing Ig-gene mutation in tumour cells, these observations suggest that BL originate in GCs. A distinctive feature of BL is the presence of a reciprocal *c-myc* proto-oncogene translocation into the immunoglobulin regulatory elements on either chromosome 14 (the heavy chain locus), chromosome 22 or chromosome 2 (the light chain locus) (Manolov and Manolova 1972; Dalla-Favera *et al.,* 1987; Klein 1983). This translocation leads to over-expression of the c-myc protein contributing to the high rate of cell proliferation in BL tumours.

### 1.14.4 Hodgkin's lymphoma

Hodgkin's lymphoma is one of the most frequently occurring lymphomas in the Western world (Schmitz *et al.,* 2009). There are two types of HL, lymphocyte predominant (LP-HL) and classical (c-HL). The latter is further divided into four subtypes: nodular sclerosis (NS); mixed cellularity (MC); and the rarer forms, lymphocyte rich (LR) and lymphocyte depleted

(LD). The presence of large malignant Hodgkin's Reed-Sternberg (HRS) cells characterises classical HL. This tumour is unusual, in that only 1-2% of the total tumour mass is made up of HRS cells, with the remainder comprising infiltrating non-neoplastic cells which include B cells, T cells, macrophages, neutrophils and eosinophils. The malignant cells of LP-HL, lymphocytic and histocytic cells, differ immunophenotypically and morphologically from the malignant HR-S cells of c-HL; HR-S cells are generally multinucleated whereas L&H cells are typically mononuclear and small. HR-S cells have an unusual immunophenotype. They express markers of various haematopoietic cell lineages such as CD3, CD4 and CCL17 and show a global loss in their B cell phenotype characterised by the down-regulation of B-cell specific transcription factors such as Oct-2 and BOB1 (Kuppers 2009). It is now believed that HR-S cells are derived from pre-apoptotic germinal-centre (GC) B cells. These cells have *Ig* $V_H$ and $V_L$ gene rearrangements which are restricted to B cells. In many cases, the *Ig V*-region of the gene shows somatic hypermutations, an event that normally occurs in antigen stimulated mature B cells proliferating in GCs (Schmitz *et al.,* 2009).

An association between EBV and HL first emerged from serological studies that found elevated EBV antibody titres in HL patients (Levine Ablashi *et al.,* 1971). Subsequently, it was found that individuals who suffered from IM were at an increased risk of developing this disease (Gutensohn and Cole 1980). This was followed by the discovery of monoclonal EBV genomes in HR-S cells (Weiss *et al.,* 1989). Approximately 50% of cases of HL test positive for EBV and it has been proposed that the virus may be responsible for the rescue of pre-apoptotic cells since only EBV-positive cases of HL harbour crippled Ig gene rearrangements (Siemer *et al.,* 2008). In these cases BCR-deficient GC B cells may be rescued from apoptosis by EBV infection, possibly by LMP2A (Bechtel *et al.,* 2005). Further evidence supporting a role for EBV in the pathogenesis of HL is provided by observations from our

laboratory, which showed that the major EBV oncogene LMP1 could induce in GC B cells, the presumptive progenitors of HL, the down-regulation of B cell genes which is characteristic of this condition (Vockerodt *et al.,* 2008).

Although no global methylation analysis has been carried out on HRS cells compared to their normal counterpart, many of the B cell specific genes down-regulated in HL, for example, CD19, MS4A1, CD79B, SPI1, SYK, POU2AF1, TNFRSF17, LCK, TCL1A and, IGH@ have been shown to be silenced by DNA methylation in HL cell lines and HRS cells (Ushmorov *et al.,* 2005; Doerr *et al.,* 2005; Ushmorov *et al.,* 2004). Treatment of B cell lines with both 5-aza-deoxycytidine, a DNA demethylating agent in combination with trichostatin A, a histone acetylating agent led to the induction of a Hodgkin-like gene expression programme in the B cell lines (Ehlers *et al.,* 2008).

However, a number of important question remain unanswered. What is the impact of EBV infection on the transcriptional profile of GC B cells? Is EBV infection of these cells followed by changes in the expression of the DNMTs. What consequences, if any, does this have for the cellular epigenome? Is the expression of the DNMTs de-regulated in HL. I aim to address these questions in my thesis.

## 1.15 DNMT inhibitors in EBV-associated malignancies

As methylation plays a major role in the silencing of EBV promoters, a demethylating agent may re-activate those promoters which could in turn facilitate immune mediated destruction of tumour cells. A clinical trial has been undertaken in which biopsies from 10 patients with EBV- associated malignancies were taken before and after treatment with 5-azacytidine (Chan *et al*, 2004). Although analyses of several EBV promoters before and after treatment in patients with NPC or AIDS lymphoma revealed demethylation of the latent promoters Cp, Wp and LMP1 and also at the lytic promoters BZLF1 and BRLF1, the only viral gene to be reactivated was BZLF1. However, it has been suggested that HDAC inhibitor treatment in combination with the DNMT inhibitor may lead to the re-expression of other silenced genes (Chan *et al,* 2004). A phase 2 clinical trial of azacitidine and a HDAC inhibitor is currently being carried out on patients with relapsed or refractory Hodgkin lymphoma. Therefore there are compelling reasons for attempting to understand in more detail the methylation changes in cellular genes induced by EBV. This is one of the objectives of my thesis.

Chapter 2

# Materials and Methods

## 2.1 Cell culture

### 2.1.1 Maintenance of cell lines

Suspension cells were maintained in exponential growth at 37ºC in 75 $cm^3$ flasks in RPMI 1640 medium (Sigma) supplemented with 10% v/v selected foetal calf serum (FCS; PAA the cell culture company) and 1% v/v penicillin/streptomycin (Gibco). Adherent cell lines were maintained in 10 cm dishes in (N-[2-Hydroxyethyl] piperazine-N'-[2-ethanesulphonic acid]) (HEPES)-buffered Dulbecco's Modified Eagle's Medium (DME/HEPES, Sigma) with the same supplements as above. Once cells were approximately 80% confluent, they were sub-cultured to reduce cell density and prevent inhibition of growth.

#### 2.1.1.1 Culturing non-adherent cells

For non-adherent cells, media was removed from flasks and centrifuged. The supernatant was discarded and cell pellets re-suspended in fresh pre-warmed media, counted and the appropriate number of cells transferred into sterile tissue culture vessels.

#### 2.1.1.2 Culturing adherent cells

Prior to splitting, adherent cells were washed twice in phosphate-buffered saline (PBS: 1 PBS tablet dissolved in 100 ml of water, sterilised by autoclaving at 121ºC, 15 psi for 15 minutes). Trypsin solution (Gibco) was then added to the cells and incubated at 37ºC for 5 minutes to allow detachment. The trypsinized cells were centrifuged at 1000 rpm for 5 minutes, re-suspended in media, counted, and the appropriate number of cells transferred into sterile tissue culture vessels.

### 2.1.2 Cell lines used

| Cell line | Growth Characteristic | Info (origin) | Growth conditions |
|---|---|---|---|
| GC B cell derived LCL | suspension | GC B cells infected with wild type 2089 EBV | RPMI 1640, 2mM glutamine, 10% FCS, 1% pen/strep. 37°C in 5% carbon dioxide |
| L591 | suspension | EBV-positive cell line that originated from the pleural effusion of a female patient with histologically confirmed nodular sclerosis HL (Diehl *et al.*, 1982). | RPMI 1640, 2mM glutamine, 10% FCS, 1% pen/strep. 37°C in 5% carbon dioxide |
| L1236 | suspension | EBV-negative cell line established from the peripheral blood of a patient with advanced HL (Wolf *et al.*, 1996). The HRS cell origin of L1236 cells has been confirmed by the finding of identical Ig gene rearrangement sequences in L1236 cells and HRS cells of the same patient's bone marrow (Kanzler *et al.*, 1996). | RPMI 1640, 2mM glutamine, 10% FCS, 1% pen/strep. 37°C in 5% carbon dioxide |
| L540 | suspension | EBV-negative HL cell line derived from the pleural effusion of a HL patient, and genomic analysis of the cell line has revealed monoclonal rearrangements of T-cell receptor beta and gamma loci and germ line configuration of Ig genes, suggesting a T-cell origin for the cell line (Falk *et al.*, 1987; Drexler *et al.*, 1988). | RPMI 1640, 2mM glutamine, 10% FCS, 1% pen/strep. 37°C in 5% carbon dioxide |
| L428 | suspension | EBV-negative cell line established from the pleural effusion of a patient with histologically confirmed HL (Schaadt *et al.*, 1980). | RPMI 1640, 2mM glutamine, 10% FCS, 1% pen/strep. 37°C in 5% carbon dioxide |
| KM-H2 | suspension | EBV-negative cell line originally established from the pleural effusion of a patient with mixed cellularity HL (Kamesaki *et al.*, 1986). | RPMI 1640, 2mM glutamine, 10% FCS, 1% pen/strep. 37°C in 5% carbon dioxide |
| B95.8 LCL | suspension | Marmoset monkey peripheral blood leukocytes infected with EBV (Miller and Lipman 1973) | RPMI 1640, 2mM glutamine, 10% FCS, 1% pen/strep. 37°C in 5% carbon dioxide |
| 2W LCL | suspension | LCL containing recombinant EBVs with 2 copies of the *Bam*HI W repeat generated by Dr. Rosemary Tierney (CRUK) | RPMI 1640, 2mM glutamine, 10% FCS, 1% pen/strep. 37°C in 5% carbon dioxide |
| 11W LCL | suspension | LCL containing recombinant EBVs with 11 copies of the *Bam*HI W repeat | RPMI 1640, 2mM glutamine, 10% |

| | | | |
|---|---|---|---|
| | | generated by Dr. Rosemary Tierney (CRUK) | FCS, 1% pen/strep. 37°C in 5% carbon dioxide |
| Rael | suspension | Latency I Burkitt's lymphoma line derived from endemic EBV positive BL biopsies | RPMI 1640, 2mM glutamine, 10% FCS, 1% pen/strep. 37°C in 5% carbon dioxide |
| X50-7 LCL | suspension | EBV positive LCL derived from cord blood cells. Viral genome is deleted for Cp thus they are Wp using (Miller *et al.*, 1984). | RPMI 1640, 2mM glutamine, 10% FCS, 1% pen/strep. 37°C in 5% carbon dioxide |
| HeLa | Adherent | HPV 18 positive cell line derived from cervical cancer cells | DMEM, 10% FCS, 1% pen/strep. 37°C in 5% carbon dioxide |
| 293/2089 | Adherent | 293 cells stably carrying the 2089 EBV genome. (Feederle *et al.,* 2005) | DMEM, 10% FCS, 1% pen/strep. 37°C in 5% carbon dioxide |
| Fibroblasts | Adherent | Cells originating from connective tissue. Used as a feeder layer for GC B cells. | DMEM, 10% FCS, 1% pen/strep. 37°C in 5% carbon dioxide |

### 2.1.3 Cryopreservation of cells

1 x $10^7$ cells were pelleted by centrifugation at 1000 rpm in a 5810R Eppendorf centrifuge for 5 minutes and re-suspended in 1 ml freezing solution (50% v/v supplemented medium, 40% v/v FCS and 10% v/v dimethyl sulphoxide (DMSO; Sigma)). Cells were transferred to cryopreservation tubes and cooled slowly overnight to -80ºC in a freezing box surrounded by sponge soaked in isopropanol. Cells were then moved to the vapour phase of a liquid nitrogen freezer for long-term storage.

### 2.1.4 Recovery of cells from nitrogen storage

Cells were thawed quickly at 37ºC to minimize exposure to DMSO and transferred to a 15 ml tube. Warmed RPMI with supplements was added gradually to the cells to dilute the DMSO and allow the cells to recover slowly. Cells were pelleted by centrifugation at 1000 rpm in a

5810R Eppendorf centrifuge for 5 minutes, re-suspended in 10 ml medium and transferred to a 25 $cm^3$ at 37ºC.

### 2.1.5 Cell counting

After harvesting cells in the same manner as for sub-culture, disposable Glasstic™ slides (Hycor Biomedical Ltd., Edinburgh, UK) were used to determine cell concentration in culture. 100 µl of cell suspension was removed from culture and mixed with 100 µl of trypan blue. 20 µl of this was pipetted into a well on the Glasstic™ slide. Using an inverted light microscope, all live (bright spherical) cells contained within three diagonal zones of the Glasstic™ grid were counted. This number was divided by three to obtain an average count of cell number per zone. This average value was then multiplied by $10^4$ to obtain the number of cells per ml of culture medium.

### 2.1.6 Mycoplasma testing

All cell lines were periodically tested using the MycoAlert mycoplasma detection kit and were found to be consistently negative for mycoplasma.

## 2.2 RNA, DNA and protein isolation and quantification from cells

### 2.2.1 RNA isolation

Total RNA was isolated from cells using the QIAGEN RNeasy™ mini kit according to the manufacturer's protocol (QIAGEN): 5 x $10^6$ cells were transferred directly from culture to a sterile universal tube and pelleted by centrifugation at room temperature for 5 minutes at 400 *g*. The supernatant was aspirated but the cells were not washed (to reduce the possibility of mRNA degradation). Cells were lysed using 350 µl of the lysis buffer and then incubated at room temperature for at least 5 minutes. Cell lysates were mixed thoroughly with 350 µl of 70% ethanol diluted in DNase/RNase-free water and transferred to spin columns placed in collection tubes. Columns were then centrifuged at 8000 *g* for 15 seconds. Flow-through was discarded and 700 µl Buffer RW was pipetted into spin columns before being centrifuged again at 8000 *g* for 15 seconds. Flow-through was discarded and columns were transferred to fresh collection tubes. 500 µl of Buffer RPE was pipetted into spin columns before being centrifuged at 8000 *g* for 15 seconds. Flow-through was discarded and another 500 µl Buffer RPE was added. Columns were centrifuged at 8000 *g* for 2 minutes. Flow-through and collection tubes were discarded and columns were transferred to microcentrifuge tubes with their lids removed. They were centrifuged for 1 minute at 16000 *g* to completely dry the membrane in the columns. Columns were transferred to centrifuge tubes provided in the RNeasy™ kit and 30 µl of DEPC water was pipetted directly onto membranes. Columns were centrifuged for 1 minute at 8000 *g* to elute the RNA.

### 2.2.2 RNA quantification

The RNA concentration of each sample was measured on a NanoDrop ND-1000 Spectrophotometer (Labtech International): 2 µl of DNase/RNase-free water was applied to the sample well to initialize the NanoDrop. The sample well was wiped clean with tissue; 1 µl RNase-free water (from RNeasy™ kit) was applied and used as a blank measurement. Again the sample well was wiped clean and 1 µl of eluted RNA was used to obtain RNA concentration and quality (ratio A260/280) for each sample, the well being wiped clean between each measurement.

### 2.2.3 DNA extraction from cells

#### 2.2.3.1 Cell lysis

Total cellular DNA was extracted from cells by lysis using proteinase K. 1 x $10^7$ cells were harvested by centrifugation, washed in PBS and re-suspended in 200 µl lysis buffer (1X PCR buffer (Sigma), 0.5% Tween 20 (Sigma) and 1.3 M proteinase K (Roche). Cells were incubated overnight at 55ºC.

#### 2.2.3.2 Chloroform/phenol extraction

DNA was purified by phenol/chloroform (Sigma) extractions. An equal volume of phenol was added to the sample and vortexed thoroughly. The sample was centrifuged at maximum speed in an Eppendorf microfuge for 5 minutes. The upper aqueous layer was retained and transferred to a new tube. An equal volume of phenol/chloroform was added to the sample, vortexed and centrifuged for 5 minutes. The upper aqueous layer was retained and transferred to a new tube. An equal volume of chloroform was added to the sample, vortexed and

centrifuged for 5 minutes. The upper aqueous layer was removed and precipitated in cold ethanol.

#### 2.2.3.3 DNA ethanol precipitation

The DNA was precipitated by adding 0.1X volume of 3 M sodium acetate (Sigma), 2.5X volume of 100% ethanol and 1 µl of glycogen (Invitrogen) to the aqueous layer and incubated at -20ºC overnight. DNA was pelleted by centrifugation at maximum speed in a microfuge for 30 minutes. The pellet was washed in 70% v/v cold ethanol, air dried and re-suspended in 50 µl sterile distilled water.

### 2.2.4 DNA quantification

The DNA concentration of each sample was measured on a NanoDrop ND-1000 Spectrophotometer (Labtech International): 2 µl of DNase/RNase-free water was applied to the sample well to initialize and blank the NanoDrop. The sample well was wiped clean and 1 µl DNA was used to obtain the DNA concentration and quality (ratio 260/280) for each sample. The ratio of optical density (OD) at 260 nm to 280 nm should lie between 1.8 and 2.2 for samples without contamination.

### 2.2.5 Protein extraction

1 x $10^7$ cells were harvested by centrifugation at 1000 rpm for 5 minutes. The culture media was decanted and discarded and the cell pellet was re-suspended in 5 ml PBS. Cells were washed and centrifuged again in PBS to ensure complete removal of culture media. The supernatant was discarded and the pellet re-suspended in 100 µl RadioImmuno Precipitation Assay (RIPA) buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 1% Triton X-100, 0.5% sodium Deoxycholate, 0.1% SDS, 1 mM sodium vanadate and 1X protease inhibitor cocktail

(Roche)). Samples were left on ice for 30 minutes, followed by centrifugation at 4ºC for 15 minutes. The supernatant was transferred to a new 1.5 ml microcentrifuge tube and kept on ice during the determination of protein concentration.

### 2.2.6 Determination of protein concentration

Protein was quantified using the Bio-Rad Protein Assay Kit (Bio-Rad Laboratories). 10 mg/ml stock bovine serum albumin (BSA; Sigma-Aldrich) was diluted to 1.5, 0.8, 0.4, 0.2, 0.1 and 0 mg/ml concentrations with distilled water for protein standards. 2 µl of protein samples were diluted in 18 µl distilled water. 10 µl of each sample and standard was plated out in duplicate into each well of a 96 well plate (IWAKI). Bio-Rad Protein Assay Reagent was diluted 1:5 in distilled water and 200 µl added to each standard and sample. This was incubated for 5 minutes at room temperature. The absorbance of standards was read on a Bio-Rad 680 microplate reader at 750 nm and used to plot a calibration curve from which the protein content of the samples was determined.

## 2.3 Complimentary DNA synthesis and polymerase chain reaction

### 2.3.1 Primer design

Polymerase chain reaction (PCR) primers were designed using primer 3 software (www.primer3.com). Human genomic sequences were obtained from the UCSC genome browser (www.ucsc.com). Viral genomic sequences were obtained from the EMBL-EBI website (www.ebi.ac.uk/genomes/virus). Desirable characteristics of the primer sequences were a length of 18-25 bp, G+C content of 40-60% and with an annealing temperature of 58 - 62ºC. For the amplification of complimentary DNA (cDNA), oligonucleotide primers were designed to bind specifically the gene of interest. Basic local alignment search tool (BLAST) searches (http://www.ncbi.nlm.nih.gov/BLAST/) were performed in all cases to ensure that primers were not complimentary to other regions of the genome and bound specifically to the target sequences. Primers were synthesised by Alta-Biosciences (University of Birmingham) and supplied as lyophilized pellets. Primers were reconstituted in DEPC treated water (Applied Biosystems) to a concentration of 100 µM and stocks stored at -20ºC in aliquots.

### 2.3.2 Reverse transcription (RT) reaction

RNA was reverse transcribed to cDNA using Superscript® III first-strand synthesis system (Invitrogen). 500 ng RNA from each sample was pipetted into sterile 0.2 ml PCR tubes along with 50 ng random primers (Promega), 1 µl of 10 mM deoxyribonucleotide triphosphates (Roche) and the volume made up to 10 µl using DNase/RNase-free water (Promega). A negative control with DNase/RNase-free water instead of cDNA was set up in parallel. The mixture was then incubated at 65°C for 5 minutes and then cooled on ice for 1 minute to denature RNA and prohibit re-annealing. Meanwhile, an RT-reaction master mix consisting of

the following components was made up in a sterile microcentrifuge tube for each of the RNA samples plus one (to allow for pipetting error): 2 µl 10 X RT buffer, 4 µl 25 mM $MgCl_2$, 2 µl 0.1 M DTT, 1 µl RNaseOUT™, 1 µl Superscript™ III RT (all Invitrogen). 10 µl of the master mix was added to each cDNA sample. Using an Eppendorf Thermal Cycler, cDNA was synthesized by incubating samples at 25°C for 10 minutes followed by 50°C for 1 hour. The reaction was terminated by heating the samples to 85°C for 5 minutes and then by chilling to 4°C. The cDNA samples were stored at -20°C until required.

### 2.3.3 DNA PCR and RT-PCR

A PCR-reaction master mix consisting of the following components was made up in a sterile Eppendorf tube for each of the cDNA or DNA samples plus one (to allow for pipetting error): 12.5 µl 2X PCR master mix (Promega), 2.5 µl 3' primer, 2.5 µl 5' primer (final concentration 2 µM of each, see individual chapters for specific primer sequences). 17.5 µl of this master mix was pipetted into sterile 0.2 ml PCR tubes, 2 µl of synthesized cDNA or 50 ng of DNA was added to the master mix and the total volume made up to 25 µl with RNase/DNase free water. PCR amplification in an Eppendorf Thermal Cycler involved an initial 5 minute denaturation at 94°C, followed by 30 cycles consisting of a denaturing step for 30 seconds at 94°C, an annealing step for 1 minute at a temperature specific for primers (see individual chapters), and an extension for 1 minute at 72°C. A final extension step of 72°C for 10 minutes ensures all single stranded molecules have been replicated. PCR products were stored at 4°C until required.

### 2.3.4 Agarose gel electrophoresis of PCR products

#### 2.3.4.1 Agarose gel preparation

Amplified samples were analysed by electrophoresis through agarose gels. A 1.5% agarose gel was made by melting 1.7 g agarose (Eurogentec) in 115 ml of 1X tris-borate buffer solution (TBE; 45 mM tris, 45 mM boric acid, 1 mM EDTA pH 8.3 (all Fisher Scientific)) in a conical flask. The mixture was boiled using a microwave set at full power until all agarose had dissolved. The melted solution was cooled by running cold tap water over the base of the conical flask. At this point, 3 µl of 5 mg/ml stock ethidium bromide (Fluka Biochemika) was added and the gel was poured into a standard gel casting tray (Fisher Scientific), sealed by masking tape, producing a 1 cm thick gel. A 16-tooth comb was positioned near the top of the gel plate so that 16 complete wells formed once the gel had set. The gel was allowed to stand at room temperature for 1 hour or at 4$^{o}$C for 30 minutes to harden. Once the gel had set, the comb and masking tape were removed and the gel submerged into 1X TBE in a submarine mini-gel electrophoresis unit (Fisher Scientific).

#### 2.3.4.2 Gel electrophoresis

PCR products were pulse centrifuged before use. Each 25 µl of PCR sample was mixed with 5 µl of 6X blue/orange loading dye. 5 µl of 100 bp Ready-Load™ DNA ladder (Invitrogen) was added to the first lane of the submerged gel in the gel electrophoresis tank. 30 µl of each PCR product was added to other lanes alongside a negative (water) control. The gel was run for 1 hour at 120 V, 95 mA. The negatively charged products migrate through the gel to the

cathode, separating on the basis of size. Photographs were taken under UV light and documented using a GeneFlash Syngene Bio Imaging analyzer.

## 2.4 Quantitative PCR (Q-PCR)

Q-PCR requires a combination of two oligonucleotide primers and a dual-labeled fluorogenic hybridization probe to detect the gene of interest (Heid *et al*., 1996). One fluorescent dye serves as a reporter (FAM), and its emission spectra is quenched by a second fluorescent dye (TAMRA). During the extension phase of PCR, the 5' to 3' exonuclease activity of the Taq DNA polymerase cleaves the reporter from the probe, thus releasing it from the quencher, resulting in an increase in fluorescent emission at 518 nm.

### 2.4.1 Q-PCR and Q RT-PCR reaction

All equipment was irradiated under UV light to remove any DNA contaminations. cDNA or DNA for use in the Q-PCR was prepared as described in Section 2.3.3 or 2.2.3. PCR reactions were set up on ice and performed in 96-well reaction plates (Applied Biosystems). 2 µl cDNA or 50 ng DNA was generally used per reaction; however for genes with low transcript levels more cDNA was required. 15 µl of a mastermix comprising 12.5 µl 2X Taqman universal mastermix, 1.25 µl 20X endogenous control and 1.25 µl 20X primer and probe mix (all Applied Biosystems) was added to each well along with the cDNA or DNA and the total volume made up to 25 µl with DEPC treated water (Applied Biosystems). All samples were run in triplicate and wells sealed using optically clear caps (Applied Biosystems). The plate was then placed in the ABI Prism 7700 Sequence Detection System (Applied Biosystems) and subjected to an enzyme activation step at 50ºC for 2 minutes, a denaturation step at 95°C for 10 minutes followed by 40 cycles of denaturation at 95°C for 15

seconds and an extension step at 60°C for 1 minute. Taqman assays used within the thesis are shown in Table 2.1.

### 2.4.2 cDNA synthesis for EBV RNA

The synthesis of viral cDNA was performed using a different method. Briefly, 400 ng RNA was denatured at 90ºC for 2 minutes and cooled on ice. Denatured RNA was reverse-transcribed in a 20 µl reaction containing 1X avian myeloblastosis virus reverse transcriptase (AMV-RT) reaction buffer (50 mM Tris-HCL pH 8.5, 30 mM KCL, 8 mM $MgCl_2$, 1 mM dithiothreitol (DTT) (Roche), 200 µM each of dATP, dCTP, dGTP, dTTP (Roche), 1 µM of the 3' gene-specific primer (Dr. Andy Bell), 5 units of AMV-RT enzyme (Roche) and 1 unit of RNAse inhibitor (Roche). This mixture was incubated at 42ºC for 90 minutes, followed by heat inactivation of the AMV-RT at 90ºC for 5 minutes. cDNA was diluted to a final volume of 80 µl and stored at -20ºC.

### 2.4.3 Assays to detect EBV mRNAs

Taqman Q RT-PCR was used to detect expression of EBV mRNAs. Primer/probe combinations specific for detection of Wp- and Cp- initiated, EBNA2, $Y_3$-U-K and Q-U-K spliced EBNA1, LMP1, LMP2A and BZLF1 transcripts were kindly provided by Dr. Andy Bell. All EBV specific probes were FAM-labelled, while GAPDH mRNA was quantified using a pre-developed assay reagent containing primers and a VIC-labelled probe (Applied Biosystems). 25 ng cDNA was used in a 25 µl reaction containing 1X Taqman Universal PCR Mastermix, 5' and 3' primers and probes and 1X GAPDH mix. All samples were analysed in triplicate and subjected to an initial uracil-N glycosylase incubation (50ºC for 2 minutes) and AmpliTaq Gold activation (95ºC for 10 minutes), followed by 40 cycles of 95ºC

for 15 seconds and 60ºC for 1 minute. Fluorescent signals were detected by an ABI Prism 7700 Sequence Detection System (Applied Biosystems). Template-negative and RT-negative reactions served as controls for contamination.

### 2.4.4 The comparative Ct method (ddCt) for relative quantitation of gene expression

The delta-delta (dd)Ct method (Livak and Schmittgen, 2001) enables relative quantitation of transcripts without the need for standard curves when looking at expression levels of a target gene relative to an endogenous control (e.g. GAPDH). The ddCt method was used to measure differences in the amount of target gene transcripts between samples. First, the difference (dCt) between the Ct values of the target gene and the endogenous gene was calculated for each sample studied (dCt = target Ct - endogenous Ct). The sample selected as the "baseline" sample for expression of the target gene is referred to as the reference sample. The difference between each of the samples dCt and the reference samples dCt was calculated, generating the ddCt value for each sample (ddCt = reference dCt - target dCt). The ddCt for each sample was then converted to an absolute value using the following equation: fold change in expression level = $2^{-ddCt}$.

| NCBI gene symbol | Applied Biosystems assay number |
|---|---|
| DNMT1 | Hs00154749_m1 |
| DNMT3B | Hs01027166_m1 |
| DNMT3A | Hs00171876_m1 |
| PRMT1 | Hs01587651_g1 |
| PRMT5 | Hs01047356_m1 |
| CARM1 | Hs00406354_m1 |
| PADI4 | Hs00202612_m1 |

**Table 2.1: Taqman Q-RT-PCR assays**

## 2.5 Analysis of protein expression by Western blotting

### 2.5.1 Sodium dodecyl sulphate polyacrylamide gel electrophoresis

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) gels were set up in two phases using the mini-Protean 3 Bio-rad apparatus. An 8% resolving gel containing 30:1 acylamide:bisacrylamide (Bio-rad), 390 mM Tris-HCL pH 8.8, 0.1% w/v SDS, 0.06% w/v N,N,N',N'- tetramethylethylenediamine (TEMED, Sigma) and 0.1% w/v ammonium persulphate (APS, Sigma) was made. 300 µl isopropanol was applied on top to give a uniform surface. A stacking gel containing 3% acrylamide, 125 mM Tris-HCL pH 6.8, 0.1% w/v SDS 0.1% w/v TEMED and 0.1% w/v APS was poured on top of the resolving gel, a comb was placed in the stacking gel and allowed to set for 30 minutes. The appropriate protein concentrations of samples were boiled in 2X laemilli buffer (Sigma) for 10 minutes and protein was loaded into each well together with 10 µl of a protein marker (Bio-rad). Gels were run in 1X running buffer consisting of 30 g TRIS, 144 g glycine (Fisher), 10 g SDS in 10 L distilled water. Solubilised proteins were separated by electrophoresis at 125 V, 400 mA for 2 hours.

### 2.5.2 Protein transfer

Six sponges, six pieces of Whatman® chromatography paper (Sigma-Aldrich) and one piece of BioTrace NT membrane (VWR International), all cut to gel size, were soaked in a tank of transfer buffer (30 g tris, 144 g glycine, 2 L methanol in 8 L distilled water). The 'transfer sandwich' was set up in a XCell II Blot Module (Invitrogen), ensuring air bubbles were pushed out between each layer, in the following order: 3 sponges, 3 pieces of filter paper, gel, BioTrace NT membrane, 3 pieces of filter paper, and 3 sponges. The transfer module was

then placed into the XCell *Surelock* Mini-Cell (Invitrogen) and filled with transfer buffer. Transfer buffer was poured into the tank around the plates to a depth of 10 cm. Protein was then transferred from the gel to the BioTrace NT membrane on ice at 40 V, 400 mA for 2 hours.

### 2.5.3 Protein detection

Non-specific protein binding was blocked by incubating the membrane for 1 hour at room temperature in 5% milk dissolved in tween-tris buffered saline (T-TBS), consisting of 1.21 g TRIS and 8.77 g NaCl in 1 L distilled water (pH adjusted to 7.6 using concentrated HCl) with 0.5 ml Tween 20 (Fisher Scientific). The membrane was incubated overnight at 4 ºC in diluted primary antibody (Table 2.2).

| Target protein | Antibody | Species | Dilution |
|---|---|---|---|
| DNMT1 190 kDa | Ab1-6632 | Rabbit polyclonal | 1:500 |
| DNMT1 190 kDa | IMG 07703648959 | Mouse monoclonal | 1:500 |
| DNMT3A 85 kDa | Ab-16704 | Rabbit polyclonal | 1:200 |
| DNMT3A 100-130 kDa | IMG-268A | Mouse monoclonal | 1:100 |
| DNMT3A 120 kDa | Sc-10231 | Goat polyclonal | 1:100 |
| DNMT3B 100 kDa | IMG-184A | Mouse monoclonal | 1:200 |
| DNMT3B 130 kDa | Ab-2851 | Rabbit polyclonal | 1:500 |
| PRMT1 42 kDa | Sc-59648 | Mouse monoclonal | 1:2000 |
| PRMT5 70 kDa | Ab-12191 | mouse | 1:2000 |
| CARM1 63 kDa | Ab-50278 | mouse | 1:1000 |
| HA tag | 16B12 | Mouse monoclonal | 1:1000 |

**Table 2.2: List of primary antibodies**

After rinsing three times for 5 minute in T-TBS the membrane was incubated for 1 hour in HRP-conjugated secondary IgG (DakoCytomation) and then rinsed as before.

### 2.5.4 Visualization

Antibody-protein complexes were detected using the enhanced chemiluminescence (ECL) kit (Amersham Biosciences). 1 ml of each ECL reagent was mixed in a universal tube immediately prior to use. The membrane was incubated in the ECL mixture for 1 minute before being wrapped carefully in Saran wrap (Fisher Scientific).

### 2.5.5 Autoradiography

The wrapped BioTrace NT membrane was placed into a Hypercassette™ autoradiography cassette with a sheet of Hyperfilm™ (both Amersham Biosciences) on top, which was developed for between 30 seconds to 1 hour. The Hyperfilm™ was developed in a Kodak X-OMAT 1000 processor (Kodak Limited).

## 2.6 EBV infection experiments

### 2.6.1 Preparation of virus stocks

293 cells carrying the EBV 2089 genome (Dr. Claire Shannon-Lowe) were plated into 6 well plates 24 hours prior to transfection. 0.75 µg BZLF1 expression plasmid and 0.75 µg BALF4 (gp110) plasmid (Dr. Claire Shannon-Lowe) were co-transfected per well using Lipofectamine™ (Invitrogen) transfection reagent. For each well, 750 ng of plasmid was diluted in 100µl Optimem (Gibco) and mixed gently. 1.5 µl Plus reagent was added to DNA, mixed gently and incubated for 15 minutes at room temperature. 6 µl Lipofectamine LTX reagent was added to the DNA and incubated for 30 minutes at room temperature. 100 µl of DNA-Lipofectamine LTX complexes was added to each well and mixed by rocking. Supernatant was harvested 72 hours later, centrifuged at 2000 rpm for 5 minutes and filtered using a 0.2 µm filter to remove cellular contaminants. This supernatant was used as a virus source; 100 µl was retained for quantification, the rest was stored in aliquots at -80ºC

### 2.6.2 Quantification of virus titre

An equal volume of lysis buffer (100 µg/ml proteinase K (Roche), 10 mM Tris-HCL pH 8.8, 1.5 mM $MgCl_2$, 50 mM KCL, 0.1% v/v Triton X-100) was added to the viral supernatant and incubated at 55ºC for 1 hour followed by 99ºC for 10 minutes to heat inactivate the proteinase K. Q-PCR was then carried out for the EBV polymerase gene (BALF5). 5 µl of processed viral supernatant was added per well of a 96-well plate to determine the number of virus genomes per ml of supernatant.

A multiplex PCR reaction allowed simultaneous amplification and detection of BALF5 with a FAM-labelled probe and the endogenous beta-2-microglobulin gene (B2M) with a VIC-

labelled probe. Primers and probes were designed by Dr. Claire Shannon Lowe. 5 µl of viral supernatant, 1X Taqman Universal PCR Mastermix (Applied Biosystems), 200 nM of each EBV Pol primer, 200 nM EBV Pol probe (Eurogentec), 60 nM of each B2M primer and 100 M B2M probe (Eurogentec) was added to a PCR tube and made up to 25 µl with DEPC treated water. Template-negative controls were included in each experiment to check for PCR contamination. All samples were analysed in triplicate and subjected to the same Q-PCR conditions as described in Section 2.4.3.

Ct values are directly related to the initial amount of DNA. To generate a standard curve to calculate Pol and B2M copy numbers, I used serial dilutions prepared from Namalwa BL, which is known to have 2 integrated EBV genome copies per cell (Henderson *et al.,* 1983). Assuming one Namalwa BL cell contains 6.6pg of DNA and 2 EBV genomes, then 132 ng/µl DNA corresponds to 40.000 EBV genomes/µl. This value was used as the highest standard and serial dilutions (10,000, 2000, 200, etc) made from this. A linear standard curve was generated by plotting Ct value against $\log_{10}$ Pol copy number, allowing the number of EBV genomes in the test samples to be calculated.

## 2.7 Preparation of Germinal Centre B cells from tonsillar tissue

### 2.7.1 GC B cell extractions

Tonsillar tissue was obtained from the Children's Hospital Birmingham following routine tonsillectomy. Tonsils were kept on ice at all times during processing. RPMI supplemented with 10% FCS, 1% pen/strep and 2 mg/ml Ciproxin (Bayer) was used throughout the preparation for washing and re-suspending the cells. Tonsils were processed one at a time in a petri dish containing 10 ml media. The tonsil was pressed gently using a scalpel to release the cells, these were then aspirated to a 50 ml falcon tube avoiding clumps. The tonsil was continuously tapped and pressed until all cells were released and aspirated to falcon tubes on ice. 15 ml lymphoprep (Axis-shield) was pippeted into 50 ml falcon tubes and 35 ml of cell mixture was carefully layered on top of the lymphoprep. This was then centrifuged at 2200 rpm in an Eppendorf centrifuge at room temperature for 30 minutes without a brake. A Pasteur pipette was used to remove the live white mononuclear cells to a clean falcon tube and topped up to 50 ml with RPMI. This was then centrifuged at 1000 rpm for 10 minutes at 4ºC. The supernatant was removed and the cell pellet washed with RPMI once again. The pellet was re-suspended in 20 ml autoMACs rinsing solution (MACS®) supplemented with BSA, Pen/strep and ciproxin. 200 µl of cells was used for counting with trypan blue. 100 µl of cell suspension was diluted 1:10 with autoMACs solution and mixed with trypan blue in a 1:1 ratio. The total number of tonsillar mononuclear cells (TMCs) was calculated to be used for CD10 enrichment.

### 2.7.2 CD10 selection

CD10 expressing cells are a population of cells made up of centroblasts and centrocytes. Both subsets of cells are CD10 positive, thus can be selected for using α-CD10-phycoerythrin (PE) and α -PE microbeads.

TMCs were centrifuged at 1000 rpm for 10 minutes at 4ºC and re-suspended in 100 µl AutoMACs/$10^7$cells. The PE conjugated anti-human CD10 antibody (EBioscience) was added to the cells at a 1:50 dilution and incubated for 15 minutes at 4ºC. Cells and antibody were then washed with AutoMACs solution. This was centrifuged at 1000 rpm for 10 minutes at 4ºC. Cells were re-suspended in AutoMACs at a concentration of 80 µl/$10^7$ cells. Anti-PE microbeads diluted 1:5 (20 µl beads/$10^7$ cells) were added and incubated for 15 minutes at 4ºC. 10X volume of autoMACs buffer was added and centrifuged at 1000 rpm for 10 minutes at 4ºC. LS columns and magnet (both MACS®) were assembled and cells re-suspended so that 500 µl are used in each column. The column was rinsed once with 3 ml autoMACs buffer. 500 µl of cells and beads were added directly to each filter and allowed to flow through. The column was then washed three times with 3 ml autoMACs buffer. The columns were removed from the magnet and put into fresh 15 ml falcon tubes. 5 ml autoMACs was added to each column and a plunger used to elute the cells on ice. Cells were combined in one tube and counted using trypan blue. The yield was usually about 10% of original cells.

### 2.7.3 Setting up a fibroblast feeder layer.

Fibroblasts were harvested (Section 2.1.1), 24 hours before isolating GC B cells. Cells were irradiated and seeded at 2 x$10^4$ cells per well in a 24 well plate containing 0.5ml DMEM with added supplements. Cells were incubated overnight at 37ºC.

### 2.7.4 GC B cell infection with EBV

$CD10^+$ cells were centrifuged at 1000 rpm for 5 minutes and re-suspended in the appropriate volume of RPMI supplemented with 10% v/v FCS. Media was removed from the fibroblast feeder layer and 5 $x10^6$ $CD10^+$ cells in 1 ml media were added. Virus supernatant was thawed at 37ºC, and was added to the cells at a multiplicity of infection (MOI) of 50 and incubated overnight. A negative control of $5x10^6$ $CD10^+$ cells without virus was included. The following day, 1 ml media was removed from each well and replaced with fresh RPMI/10% v/v FCS. Each well was checked microscopically to check the efficiency of infection using the gfp tag. At each time point, three wells were harvested for RNA, DNA and protein isolation. Once cells were confluent they were transferred to $25cm^3$ flasks and continued in culture, splitting every 3-4 days.

## 2.8 Microarray expression analysis

### 2.8.1 Introduction to microarray expression analysis

The GeneChip® Human Genome U133 Plus 2.0 microarray (HG-U133 Plus 2.0; Affymetrix) allows the simultaneous analysis of expression levels of more than 47,000 transcripts and variants. The sequences from >54'000 probe set on the HG-U133 Plus 2.0 array were selected from Genbank®, dbEST, and RefSeq. The sequence clusters were created from the UniGene database (Build 133) and then refined by analysis and comparison with a number of other publicly available databases giving the most comprehensive coverage. Each probe is comprised of eleven pairs of 25-mer oligonucleotides that can hybridise to a probe set target sequence. The hybridisation of target sequences to oligonucleotide probes is detected by staining with a streptavidin-phycoerythrin conjugate and measuring the light emitted at 570 nm. The signal produced is proportional to the amount of bound target at each location on the array and reflects the transcript levels present in the original target population. A series of control genes are used to allow normalization and scaling between each array so that differences in signal intensities can be compared.

### 2.8.2 Sample preparation

The HG-U133 Plus 2.0 arrays were used to globally report the gene expression changes that followed infection of GC B cells with EBV. Total RNA was extracted from the cells (Section 2.2.1) and prepared for hybridization to the arrays using GeneChip® one-cycle target labelling and control reagents (Affymetrix).

#### 2.8.2.1 First-strand cDNA synthesis

10 µg RNA was reverse transcribed to first-strand cDNA by adding 100 pmol T7$(dT)_{24}$ primer, 2 µl poly-A controls and made up to 11 µl with DEPC-treated water. This reaction was incubated at 70ºC for 10 minutes before adding a master mix consisting of; 4 µl 5X First Strand buffer, 2 µl 0.1M DTT, and 1 µl 10 mM dNTPs. The samples were incubated at 42ºC for 2 minutes before adding 2 µl Superscript II Reverse Transcriptase and incubating them at 42ºC for 1 hour followed by at least 2 minutes on ice.

#### 2.8.2.2 Second-strand cDNA synthesis

An RNase H-mediated second-strand cDNA synthesis reaction was performed using 130 µl of a master mix containing; 30 µl 5 X Second strand buffer, 3 µl 10 mM dNTPs, 1 µl 10U/µl *E. Coli* DNA ligase, 4 µl 10U/µl *E. Coli* DNA Polymerase I, 1 µl 2U/µl *E. Coli* RNAse H and 91 µl DEPC treated water was made up and added to each of the first strand reaction tubes and incubated at 16ºC for 2 hours. 2 µl T4 DNA Polymerase was added to each tube, incubated at 16ºC for a further 2 minutes, followed by the addition of 10 µl 0.5M EDTA.

#### 2.8.2.3 Clean-up of double stranded cDNA

600 µl cDNA binding buffer was combined with the cDNA reaction mix in a 1.5 ml Eppendorf tube. The sample was applied to the cDNA cleanup spin column placed in a collection tube and centrifuged for 1 minute at full speed. The flow-through was discarded and 750 µl wash buffer was applied to the column, the sample was centrifuged again for 1 minute at full speed. The spin column was transferred to a new collection tube and

centrifuged for 5 minutes at maximum speed with the lid open. The spin column was then transferred to a new 1.5 ml Eppendorf tube and 14 µl cDNA elution buffer applied directly to the membrane. This was incubated at room temperature for 1 minute and centrifuged at full speed for 1 minute.

#### 2.8.2.4 *In vitro* transcription reaction

*In vitro* transcription (IVT) was then performed in the presence of T7 RNA polymerase, a biotinylated nucleotide analog/ribonucleotide mix for complementary RNA (cRNA) amplification and biotin labelling. 34 µl of a master mix comprising of; 4 µl 10 X IVT labelling buffer, 12 µl IVT labelling NTP mix, 4 µl IVT labelling enzyme mix and 14 µl DEPC treated water was added to 6 µl of each cDNA sample previously prepared. This was incubated over night at 37ºC.

#### 2.8.2.5 Clean-up of cRNA

The cRNA was purified to remove unincorporated NTPs, salts and enzymes. 60 µl RNAse free water was added to the IVT reaction and vortexed. 350 µl cRNA binding buffer was added and vortexed. 250 µl of 100% ethanol was then added and mixed well by pipetting. The sample was applied to the cRNA cleanup column in a 2 ml collection tube and centrifuged at full speed for 15 seconds. The flow-through was discarded and 500 µl wash buffer added to the column. This was again centrifuged, flow through discarded and 500 µl 80% ethanol added to the spin column. The sample was centrifuged for 15 seconds at full speed and transferred to a new collection tube and centrifuged for a further 5 minutes at maximum speed. The spin column was transferred to a 1.5 ml Eppendorf tube, 11 µl RNase free water was applied directly to the spin column and incubated for 1 minute at room

temperature. The sample was centrifuged at full speed for 1 minute. 10 µl RNase-free water was added to the membrane and centrifuged for 1 minute at full speed. At this point the RNA concentration and A260/A280 ratio was noted using the NanoDrop spectrophotometer.

#### 2.8.2.6 cRNA fragmentation

25 µg cRNA was fragmented in 10 µl 10X fragmentation buffer made up to 50 µl with DEPC treated water. The reaction was left at 94ºC for 35 minutes and placed on ice. An aliquot of the cRNA and fragmented cRNA was run on a 1.5% agarose gel to check the RNA quality. The rest of the fragmented cRNA was stored at -80ºC until required.

### 2.8.3 Hybridization, washing, staining and scanning

Hybridization, washing and staining was performed as an in house service provided by Dr. John Arrand according to the manufacturer's instructions. Briefly, a hybridization cocktail was prepared including the fragmented target and probe array controls. The sample was then hybridized to the array for 16 hours. Immediately following hybridization the probe array was subjected to an automated washing and staining protocol on the fluidics station. The arrays were scanned using a GeneChip Scanner 3000 using software that defines the probe cells and computes an intensity for each cell. Each complete probe array image was stored in a separate data file identified by the experiment name and was saved with a data image file (.dat) extension.

### 2.8.4 data analysed using R and Microsoft Excel software packages.

The analysis of these arrays was performed with the help of Dr. Wenbin Wei. Images of the microarrays were analysed using Affymetrix microarray Suite 5.0. Probe level quantile

normalization and robust multiarray analysis on the raw .CEL files were performed using the Affymetrix package of the Bioconductor (http://www.bioconductor.org) project. Differentially expressed genes were identified using Significance Analysis of Microarrays (SAM).

## 2.9 Methylation Arrays

### 2.9.1 Introduction to GeneChip® Human Promoter 1.0R Array

The Affymetrix human promoter array 1.0R is designed for genome-wide studies of transcription factor binding sites, DNA methylation, histone protein modifications or other chromatin-protein interactions. It is a single array comprised of over 4.6 million probes tiled through over 25,500 human promoter regions. Probes are tiled at an average resolution of 35 bp, as measured from the central position of adjacent 25-mer oligos, leaving a gap of approximately 7.5 kb upstream through 2.45 kb downstream of the 5' transcription start sites. Sequences used in the design of the human promoter array were selected from NCBI human genome assembly (Build 34) and the promoter regions were selected using sequence information from; Refseq, ENSMBL and GenBank.

### 2.9.2 Preparation and sonication of genomic DNA

DNA was isolated from LCL and GC B cell samples (section 2.2.3). 10 µg of DNA diluted in 400 µl TE buffer was sheared to 300 - 1000 bp by sonicating the DNA for 15 minutes with 30 second pulses followed by 30 second incubations on ice using a Diagenode® sonicator. 20 µl was removed and added to 4 µl 6 X orange/blue loading dye and run on a 2% agarose gel to ensure the fragments were the correct size. The sheared DNA was then ethanol precipitated (Section 2.2.3.3). The DNA was resuspended in 30 µl water and measured on the NanoDrop.

### 2.9.3 Methylated DNA immunoprecipitation

6 µg sheared DNA was made up to 450 µl TE buffer, denatured at 95ºC for 10 minutes and then cooled on ice for 10 minutes. 51 µl 10 X IP buffer (100 mM Na-Phosphate pH 7.0, 1.4 M NaCl and 0.5 % Triton X-100) and 10 µg 5-methylcytosine antibody (Eurogentec) was added and incubated overnight at 4ºC on a rotating wheel. 40 µl sheep anti-mouse IgG Dynabeads (Dynal Biotech) were washed twice in 1 ml PBS and collected each time using a magnetic rack (Dynal Biotech). The beads were resuspended in 40 µl 1 x IP buffer and added to the antibody/DNA sample which was left for 6 hours at 4ºC on a rotating wheel. Beads were collected using the magnetic rack, supernatant was removed and the unbound fraction stored before being purified with the elution fraction. The beads were washed three times for 5 minutes in 1 x IP buffer. The beads were collected for the final time and resuspended in 250 µl proteinase K digestion buffer (50 mM Tris pH 8.0, 10 mM EDTA, 0.5 % SDS). 4 µl 20 mg/ml proteinase K (Roche) was added to the beads and incubated over night at 50ºC at 1000 rpm. The beads were captured and a phenol/chloroform extraction and ethanol precipitation was carried out on the supernatant (Section 2.2.3.2). The DNA was re-suspended in 30 µl TE buffer.

### 2.9.4 Checking for enrichment of methylated DNA by Q-PCR

Enrichment was measured in triplicate by Q-PCR using primers specific for the methylated XIST sequence, and the unmethylated GAPDH sequence. Q-PCR was performed using 2.5 µl of isolated DNA, 12.5 µl 2X SYBR green mastermix (Applied Biosystems), 10 pmol forward and reverse primer in a total volume of 25 µl made up with DEPC treated water. Amplification involved an enzyme activation step at 50ºC for 2 minutes, a denaturation step

at 95°C for 15 minutes followed by 40 cycles of denaturation at 95°C for 15 seconds and an extension step at 60°C for 1 minute. Q-PCR reactions were also performed on input DNA at 10 ng/µl, 1 ng/µl, 0.1 ng/µl, 0.01 ng/µl concentrations to generate a standard curve. The Ct versus log DNA concentration was plotted. The Ct value of the samples was used to extrapolate the DNA concentrations from the standard curve. The percent enrichment was calculated using the following equation.

$$\% \text{ enrichment} = \frac{\text{Total amount of enriched DNA}}{\text{Amount of DNA in IP}} \times 100$$

### 2.9.5 Amplification of immunoprecipitated DNA

Amplification was carried out using round A/B/C random amplification. 10 µl immunoprecipitated DNA or input DNA was added to 4 µl 5 X Sequenase reaction buffer (USB, Affmetrix) and 4 µl 200 µM primer A. DNA was heated to 95ºC for 4 minutes and cooled on ice. 0.1 µl 20 mg/ml bovine serum albumin (BSA), 1 µl 0.1 M DTT, 0.5 µl 25 mM dNTPs and 1 µl 1.3 U/µl sequenase (USB, Affymetrix) was added to the primed DNA. This mixture was ramped from 10ºC to 37ºC over 9 minutes, held at 37ºC for 8 minutes, incubated at 95ºC for 4 minutes and snap cooled on ice. 1 µl 1.3 U/µl sequenase was added to the sample, held at 10ºC for 5 minutes, ramped to 37ºC over 9 minutes, held at 37ºC for 8 minutes , incubated at 95ºC for 4 minutes and snap cooled on ice. 1 µl 1.3 U/µl sequenase was again added to the sample, held at 10ºC for 5 minutes, ramped to 37ºC over 9 minutes, held at 37ºC for 8 minutes , incubated at 95ºC for 4 minutes and snap cooled on ice . 1 µl 1.3 U/µl sequenase was added to the sample once again and the ramping and cooling process carried out once more in the same way. The sample was held at 4ºC. The samples were then purified

using a reaction cleanup kit (Qiagen) as follows; 300 µl buffer ERC was added to each reaction, pipetted into a column and centrifuged at 13'000 rpm for 1 minute. 750 µl buffer PE was added to the column and centrifuged at 13'000 rpm for 1 minute. The empty column was centrifuged at 13,000 rpm for 1 minute. 10 µl of sterile water was added to the membrane and incubated at room temperature for 5 minutes. The DNA was eluted by centrifuging the tube for 1 minute at 13'000 rpm. 20 µl of sterile water was again added to the membrane, incubated and centrifuged at 13'000 rpm for 1 minute. 30 µl eluted DNA in a 0.2 ml PCR tube, was added to 10 µl 10 X PCR buffer (Promega), 3.75 µl dNTP and dUTP mix (10 mM dCTP, dATP, dGTP, 8 mM dTTP, 2 mM dUTP), 100 µM primer B. 1 µl $MgCl_2$, 5 U/µl Taq polymerase (Promega) and 48.75 µl DEPC treated water. The mixture was subjected to PCR; 95ºC for 30 seconds, 45ºC for 30 seconds, 55ºC for 30 seconds, 72ºC for 1 minute, 15 times. The samples were then set for 25 cycles of 95ºC for 30 seconds, 45ºC for 30 seconds, 55ºC for 30 seconds, 72ºC for 1 minute plus 5 seconds for each cycle thereafter. The sample was then checked again for enrichment of methylated sequences as described in the previous section. The samples were purified using cDNA cleanup columns (Affymetrix). The cDNA was eluted in a final volume of 20 µl and measured using a NanoDrop spectrophotometer and stored at -20ºC.

### 2.9.6 Fragmentation of the samples

7.5 µg DNA was fragmented by adding 4.8 µl 10 X fragmentation buffer, 1.5 µl 10 U/µl UDG, 2.25 µl APE and made up to 48 µl with nuclease free water. This was then incubated at 37ºC for 1 hour, 93ºC for 2 minutes and held at 4ºC for at least 4 minutes. 45 µl was transferred to a new tube and 3 µl run on a 1.5% agarose gel as described in Section 2.3.4 to ensure complete fragmentation. 1 µl of DNA was also used for analysis on the bioanalyzer.

### 2.9.7 Labelling fragmented DNA

A double-stranded DNA labelling mix was prepared as follows; 12 µl 5 X TdT buffer, 2 µl TdT and 1 µl 5 mM DNA labelling reagent were added together and mixed thoroughly. 15 µl of this mix was added to the DNA sample and incubated at 37ºC for 1 hour, 70ºC for 10 minutes and held at 4ºC for at least 2 minutes.

### 2.9.8 Hybridization, washing, staining and scanning

Hybridization, washing and staining was performed as an in house service provided by Dr. John Arrand according to the manufacturer's instructions. A GeneChip Scanner 3000 was used to scan the promoter methylation arrays.

### 2.9.9 Analysis of methylation arrays

Raw array data from each of the biological replicates were quantile-normalized. TileMap software was used to identify regions of hypermethylation and hypomethylation. The output from TileMap included final summaries for each probe and a *.bed file containing selected genomic regions that could be visualized using the Affymetrix Integrated Human Genome browser (IGB).

## 2.10 Methylation Analysis

### 2.10.1 Sodium bisulphite treatment of DNA

Following sodium bisulphite treatment all unmethylated cytosines are deaminated and sulfonated, converting them to uracils, while 5-methylcytosines are protected and unchanged. Bisulphite conversion of DNA results in uncomplimentary single-stranded DNA molecules and subsequent PCR analysis results in the incorporation of thymines in place of the converted uracils. The EZ DNA Methylation-Gold Kit (Zymo research) was used to modify DNA samples according to manufacturer's instructions. Briefly, 500 ng of DNA isolated from cells as described in Section 2.2.3 was re-suspended in 20 µl DEPC treated water and added to 130 µl CT conversion reagent in a 0.2 ml PCR tube. The sample was subject to PCR; denatured at 98ºC for 10 minutes followed by an incubation at 64ºC for 2.5 hours and finally cooled to 4ºC. DNA was then applied to a Zymo-Spin™ IC Column placed in a collection tube and 600 µl binding buffer added. Sample was centrifuged at full speed for 30 seconds and flow-through discarded. The column was washed with 100 µl wash buffer and centrifuged at full speed for 30 seconds. 200 µl of desulphonation buffer was added to the column and left to stand at room temperature 20 minutes. The sample was centrifuged at full speed for 30 seconds and flow-through discarded. Column was washed twice with 100 µl wash buffer and centrifuged at full speed for 30 seconds. Finally, 10 µl elution buffer was added directly to the column and centrifuged at full speed for 30 seconds. Bisulphite modified DNA was made up to 50 µl with DEPC-treated water and stored at -20ºC for up to 6 months.

### 2.10.2 Pyrosequencing

Pyrosequencing is a method of DNA sequencing which allows quantitative analysis of CpG methylation. The method allows sequencing of a single strand of DNA by synthesizing the complementary strand along it, one base pair at a time, and detecting which base was actually added at each step. The template DNA is immobile, and solutions of A, C, G, and T nucleotides are added and removed after the reaction, sequentially. Light is produced only when the nucleotide solution complements the first unpaired base of the template. The sequence of solutions which produce chemiluminescent signals allows the determination of the sequence of the template.

#### 2.10.2.1 Pyrosequencing primer design

All primers were designed using Biotage PSQ primer design software. The genomic sequences were obtained from the UCSC genome browser (http://genome.ucsc.edu) and the sequences were then *in silico* bisulphite converted. The bisulphite converted sequence was used to generate a biotinylated and non- biotinylated primer as well as a sequencing primer under the allele quantification assay type settings. Primers were designed (www.Biomers.net.) to be no more than, between 18-24 bp in size and have an annealing temperature of ~60ºC. Products were all 100 - 200 bp in length.

#### 2.10.2.2 PCR of bisulphite modified DNA for pyrosequencing

PCR was performed in a total volume of 50 µl including 25 µl hotstart taq master mix (Thermo Scientific), 5 pmol biotinylated primer, 10 pmol non-biotinylated primer and 10 µl bisulphite modified DNA. No DNA controls were also amplified. Cycling conditions were

as follows: denaturation at 95ºC for 15 minutes, followed by 50 cycles at 95ºC for 15 seconds, 54-60 ºC for 30 seconds, 72 ºC for 30 seconds and a final extension at 72ºC for 10 minutes. Primer details and annealing temperatures are listed in Table 2.3. 10 µl from each PCR was mixed with 2 µl of 6X orange/blue loading dye (Promega) and run on a 2% agarose gel (Section 2.3.4). A strong single band was required for pyrosequencing.

#### 2.10.2.3 Pyrosequencing reaction

The pyrosequencing reactions were performed using a Pyromark ID system (Biotage) according to the manufacturer's instructions. Briefly, 40µl of the biotinylated PCR product was bound to 3 µl streptavidin beads (GE healthcare) in 37 µl binding buffer (Biotage) on a 96 well plate (Applied Biosystems) and left to shake for 5 minutes at 1300 rpm. Meanwhile, 15 pmol sequencing primer and 38.5 µl annealing buffer (Biotage) was added to each well of a sequencing plate (Biotage). Biotinylated and sequencing primer controls were also included on the sequencing plate when running a set of primers for the first time. The vacuum preparation tool was used to capture the beads from the 96 well plate. The beads were lowered into each of the following solutions for 5 seconds: 70% ethanol, denaturation buffer (Biotage) and wash buffer (Biotage). Beads were released onto the sequencing plate, heated to 80ºC for 2 minutes and left to cool. The appropriate volume of Pyrogold SQA reagents (enzyme, substrate and dNTPs; Biotage) were loaded into the reagent cartridge (Biotage) and the plate set to run. Data were analysed using the pyromark ID software.

| Primer | Sequence | Product (bp) |
|---|---|---|
| Wp region 1 fwd | GAGGGGAAAAGAGGAATAAGTT | 177 |
| Wp region 1 rev | CCCTAAAACTAACAATTAACTACT | |
| Wp region 1 seq | AGTGGGTTTGTTTGTGATT | |
| Wp region 2 fwd | TTGATATTTTAGAGTTTTGGAGGAT | 193 |
| Wp region 2 rev | TCCTCTTTTCCCCTCTAAAAATA | |
| Wp region 2 seq | CATAAACCCCCTCCTTCCTA | |
| RBM5 fwd | GAGTTTTAAGTGATAGGATGGTTT | 141 |
| RBM5 rev | ATCTCCTACCCACAATACCTCA | |
| RBM5 seq | AGTTTTAAGTGATAGGATGG | |
| ELL3 fwd | GGATTTTGAGATAGATTATGTTGA | 248 |
| ELL3 rev | TTTTCCCTTCAAACAAATTTCC | |
| ELL3 seq | GGATTTTGAGATAGATTATG | |
| MAGEA3 fwd | TTGAGAGAGGGGGAAGAGTGAGTT | 187 |
| MAGEA3 rev | TCCAACAAAACAAAAACCCACTACTAAA | |
| MAGEA3 seq | TGAGAGAGGGGGAAGA | |
| ICMT fwd | TTGAGGATTGTGAATGATTGAGT | 196 |
| ICMT rev | CCCTACAAACCCTCTAATCTATCC | |
| ICMT seq | TTTTTGGGGTTTGTGT | |
| GRB10 fwd | TGGGATGTTTTTTGTGAATTTAT | 158 |
| GRB10 rev | TTTTACTCATTCCTTACCCTAATA | |
| GRB10 seq | ATGTTTTTTGTGAATTTAT | |
| SMAD4 fwd | TTTTTTTATTTTGAGTAAAGATTAGGTTTTGT | 155 |
| SMAD4 rev | AAATATTAACATACTTAATTATAACCAAACA | |
| SMAD4 seq | TTTTTTTATTTTGAGTAAAGATTAG | |
| SPRY2 fwd | GAGAGATTTTAAGGTTTGTTTGAGTA | 218 |
| SPRY2 rev | ACCCCAATTATACCATCAACAAC | |
| SPRY2 seq | AACTACTACTAATACTTATCCTC | |
| ID2 fwd | ATGGGAGAAGGTATTGTTTTAA | 229 |
| ID2 rev | AACCAATATAAACAAACATCTTTAA | |
| ID2 seq | TGTAAAGTGTAAAGTGAAAT | |
| PRDM1 fwd | TAGGTTTGGTTTTTTATTTAGTGA | 233 |
| PRDM1 rev | CACTTTTATCAATTACCTTTTCCA | |
| PRDM1 seq | TTTTATCAATTACCTTTTCC | |
| TCL6 fwd | AGGTTAGGGTTGTTTTGTGTATAT | 259 |
| TCL6 rev | CAAACATTTCTAAAACCTCTTTAA | |
| TCL6 seq | TGTGAAGAAGATATAGTAAG | |
| CSMD1 fwd | GGGGTTGGTTTTTAAATATTGATG | 146 |
| CSMD1 rev | TTATCTTTATCAATCCTCCTTCCA | |
| CSMD1 seq | TTAAATATTGATGTGGTTAA | |
| FGFR2 fwd | GATTTGGTATTGGGGAAGATTTTT | 150 |
| FGFR2 rev | AATCCCATCTACACACTTCCTCTA | |
| FGFR2 seq | TTGGTATTGGGGAAGA | |

**Table 2.3 List of primers used for pyrosequencing**

## 2.11 Bacteriology

### 2.11.1 Preparation of L-agar/ampicillin plates

L-broth agar (LBA; 10 g L-broth powder and 7.5 g nutrient agar (both Invitrogen ) in 500 ml distilled water) and L-broth solution (10 g L-Broth powder in 500 ml distilled water) were made up in advance, sterilised by autoclaving at 121ºC, 15 psi for 15 minutes and allowed to cool for one hour. Ampicillin (Sigma-Aldrich) was added to the LBA to a final concentration of 100 µg/ml. Agar was poured into Petri dishes (Bibby Sterilin) and cooled to room temperature.

### 2.11.2 Bacterial transformation

"Top 10" chemically competent *E. coli* bacterial cells were thawed on ice. Approximately 500 ng of plasmid was added to 50 µl of the bacteria and incubated on ice for 30 minutes. Cells were heat-shocked at 40ºC for 1 minute. 200 µl of L-broth was added which were left to recover for an hour at 37ºC. 100 µl of cell suspension was spread onto L-agar/ampicillin plates and incubated inverted overnight at 37ºC.

### 2.11.3 Purification of plasmid DNA

Colonies grown on L-agar/ampicillin plates were used to inoculate 2 ml L-broth supplemented with 100 µg/ml ampicillin. Cultures grown for approximately 6 hours at 37ºC with shaking were used to inoculate 100 ml of L-broth supplemented with 100 µg/ml ampicillin and grown overnight at 37ºC with shaking. Bacterial cultures were centrifuged at 6000 g for 15 minutes at 4 ºC. Plasmid DNA was purified using a plasmid maxi prep kit (Qiagen). Briefly, bacteria was resuspended in 10 ml buffer P1 followed by vigorous mixing

and a 5 minute incubation with 10 ml buffer P2. 10 ml buffer P3 was added to the mixture and incubated on ice for 20 minutes. The suspension was centrifuged at 20,000 g for 30 minutes at 4ºC and the supernatant containing the plasmid DNA removed. The supernatant was applied to a pre-equilibrated Qiagen column and allowed to pass through. Columns were washed twice with buffer QC and DNA eluted in 15 ml buffer QF. DNA was then precipitated by adding 10.5 ml isopropanol and centrifuged at 15,000 g for 30 minutes at 4ºC. Supernatant was decanted, the pellet washed in 5 ml 70% ethanol, centrifuged at 15,000 g for 10 minutes, supernatant decanted again and the pellet allowed to air dry. DNA was resuspended in 200 µl TE buffer and the concentration determined using the NanoDrop.

### 2.11.4 Restriction

A reaction master mix was prepared using; 2 µl reaction buffer, 0.5 µl of appropriate restriction enzymes (Roche) and 15 µl DNase/RNase-free water. 18 µl of this master mix was then pipetted into 0.2 ml PCR tubes and 2 µl DNA was added to each, and left for 2 hours at 37$^{o}$ C or at room temperature overnight. The sample was run on a 1.2% agarose gel (Section 2.3.4). The gel was viewed under UV light and photographed using a GeneFlash Syngene Bio Imaging analyzer. DNA samples should have 3 bands – top is uncut DNA with vector, middle is cut vector and bottom is cut DNA insert. Vector samples should have 2 bands – top is uncut empty vector, bottom is cut vector.

## 2.12 Transfection of plasmid DNA into mammalian cells

### 2.12.1 Lipofectamine mediated transfection of HeLa cells

Plasmid DNA was transiently transfected into cells using Lipofectamine™ LTX and PLUS reagent (Invitrogen). Prior to the day of transfection, HeLa cells were plated in 500 µl of media at 6 x $10^4$ cells per well of a 24 well plate. For each well, 250 ng of plasmid was diluted in 100 µl Optimem (Gibco) and mixed gently. 0.5 µl Plus reagent was added to DNA, mixed gently and incubated for 15 minutes at room temperature. 1.25 µl Lipofectamine LTX reagent was added to the DNA and incubated for 30 minutes at room temperature. 100 µl of the DNA-Lipofectamine LTX complexes were added to each well and mixed by rocking. Media was changed the following day. Cells were harvested 48 hours following transfection.

### 2.12.2 Electroporation of suspension cells

Cells were split 24 hours before transfection to ensure they were in the log phase of growth. Cells were centrifuged at 1000 rpm for 10 minutes at 4ºC and re-suspended in 10 ml PBS. 1 x $10^7$ cells were required per reaction and each reaction was carried out in duplicate for the plasmid containing the gene of interest and for the vector control. Cells were counted and the appropriate volume of cells removed. Cells were centrifuged at 1000 rpm for 10 minutes at 4ºC and re-suspended in 10 ml Optimem (Gibco). Cells were centrifuged again and re-suspended in 300 µl Optimem per reaction. The optimised amount of DNA plasmid (10 – 30 µg) was pipetted into a 1.5 ml Eppendorf and 300 µl of cells in Optimem were added and mixed gently. The cells plus DNA were transferred to sterile electroporation cuvettes (Geneflow) and placed inside the Bio-Rad Gene Pulser. Cells were electroporated at 230 V, 975 µF and quickly transferred using a Pasteur pipette to 10 ml warm media in a 25 $cm^3$ tissue

culture flask. Cells were grown at 37ºC and the media was changed the following day. Cells were harvested 48 hours post transfection.

## 2.13 Cross-linked chromatin immunoprecipitation (X-ChIP)

Cross-linked chromatin immunoprecipitation (X-ChIP) was used to study DNMT binding *in vivo*, a method previously optimised by Dr. Laura O Neill (Institute of Infection and Immunity). Briefly, formaldehyde was used to cross-link protein-DNA interactions in the cell. Chromatin was then prepared from these cells and an immunoprecipitation (IP) performed using a specific antibody. Immune complexes were collected with protein G magnetic beads, the protein was removed and the DNA eluted and purified. DNA in the precipitates was then analysed by Q-PCR for cellular and EBV DNA.

### 2.13.1 Cross-linking and chromatin shearing

IP reactions were performed on cells split 24 hours prior to harvesting to ensure they were in log phase of growth. Cells were harvested in three times the volume of ice cold PBS containing 5 mM sodium butyrate (Millipore) and counted. Chromatin was isolated from 4 x $10^6$ cells, the appropriate volume of cells was removed and pelleted by centrifugation at 400 g at 4ºC for 5 minutes in an Eppendorf centrifuge. Cells were resuspended in 2 ml PBS/sodium butyrate. 37% formaldehyde solution (Fisher) was added to the cells to a final concentration of 1% and incubated at room temperature for 8 minutes on a roller. To stop the cross-linking reaction, glycine was added to a final concentration of 0.15 M and samples incubated at room temperature for 5 minutes. Cells were pelleted by centrifugation at 470 g at 4ºC for 10 minutes and the supernatant removed. Cells were washed twice in 2 ml ice cold PBS/sodium butyrate, centrifuged and the pellet re-suspended in 520 µl lysis buffer (10 mM EDTA, 50

mM Tris-HCL pH 8, 1% w/v SDS, 5 mM Na butyrate, 0.01 mM PMSF (Sigma), 1X protease inhibitor cocktail (Roche)). Lysate was then split into four 0.5 ml Eppendorf tubes and sonicated on ice using a Bioruptor (Diagenode) for 30 minutes at the highest power with 0.5 minute breaks in between each pulse. Samples were centrifuged at 10,000 g in a table top centrifuge at 4ºC for 10 minutes. The supernatant was removed and the DNA measured on the NanoDrop at A260/280. 2 µg of chromatin was de-crosslinked in 500 µl lysis buffer in containing 50 mM proteinase K at 68ºC for 2 hours, to check DNA size. DNA was extracted using the ChIP DNA clean and concentrator™ kit (Zymo Research). The eluted DNA in a total volume of 30 µl was run on a 1.5% agarose gel (Section 2.3.4). DNA fragments were required to be 200 – 1000 bp in size. Chromatin was then either used immediately for IP or frozen at -80ºC.

### 2.13.2 Preparation of Protein G Dynabeads

For each IP reaction, 100 µl Protein G Dynabeads (Invitrogen) were transferred to 1.5 ml Eppendorf tubes and the washing procedure carried out facilitated by the use of a magnet (Dynal MPC). The beads were placed on the magnet for 1 minute and the supernatant removed. Beads were washed three times in 0.5 ml cold RIPA buffer (10 mM TRIS-HCL pH7.5, 1 mM EDTA, 0.5 mM EGTA (Sigma), 1% Triton X-100, 0.1% SDS, 150 mM NaCl, 0.1% sodium doeoxycholate). Finally the beads were re-suspended in 90 µl RIPA buffer.

### 2.13.3 Immunoprecipitation

100 µl of beads in RIPA buffer were incubated with 2.4 µg of a polyclonal antibody over night at 4ºC with rotation. A non-specific IgG antibody and an IP with no antibody were used as controls and run with every IP. Antibodies used are shown below. After incubation, the

antibody bound beads were isolated using the magnet and the supernatant removed. 10 µg sonicated chromatin diluted in 100 µl RIPA buffer was added to the beads and incubated for 8 hours at 4ºC on a rotating wheel. The immune complexes were isolated using the magnet and washed three times with 100 µl RIPA buffer end over end for 4 minutes. The beads were re-suspended in 100 µl TE buffer pH8.

| Target protein | Antibody | Species |
|---|---|---|
| DNMT1 | Ab1-6632 | Rabbit polyclonal |
| DNMT1 | IMG 07703648959 | Mouse monoclonal |
| DNMT3A | Sc-10231 | Goat polyclonal |
| DNMT3B | Ab-2851 | Rabbit polyclonal |

**Table 2.4 Antibodies used in X-ChIP**

### 2.13.4 DNA elution and reversal of cross-links

The DNA-bead complexes were captured, TE removed and the beads re-suspended in 150 µl elution buffer (20 mM TRIS-HCL pH7.5, 5 mM EDTA, 5 mM sodium butyrate, 50 mM NaCl, 1% SDS), 50 mM proteinase K and incubated over night at 68ºC on an Eppendorf thermomixer. Beads were again captured using the magnet and the supernatant transferred to a new Eppendorf. Beads were washed in 150 µl elution buffer, supernatant removed and added to the previously collected supernatant. A further 200 µl elution buffer without SDS was added to the eluent.

### 2.13.5 DNA isolation and purification

Input and IP DNA was purified using the ChIP DNA clean and concentrator™ kit. Briefly, 5 volumes of ChIP DNA binding buffer was added to each sample, mixed and transferred to a spin column. The sample was centrifuged at 13'000 RPM for 30 seconds and the flow-

through discarded. 200 µl wash buffer to the column. The sample was centrifuged at 13'000 RPM for 30 seconds and the flow-through discarded. The wash step was repeated. The column was transferred to a new 1.5 ml microcentrifuge tube and 30 µl elution buffer added to the column and centrifuged at 13'000 RPM for 30 seconds.

### 2.13.6 Q-PCR to amplify DNA in the X-ChIP immunoprecipitates

Q-PCR was performed using SYBR green to detect DNA in chromatin IPs and input samples. Overlapping primers were designed in the Wp and Cp promoter to detect where the DNMTs were binding. GAPDH primers were designed as a negative control. The primers used are listed below. Q-PCR reactions were run in triplicate with 1 µl DNA, 12.5 µl SYBR green master mix, 15 mM forward and reverse primers in a total reaction volume of 25 µl. Cycling conditions were as follows: denaturation at 95ºC for 15 minutes, followed by 40 cycles at 95ºC for 15 seconds, 58ºC for 30 seconds, 72ºC for 1 minute. The amount of DNA present in the IP fractions was determined as a % of input and compared to the negative controls.

| Primer | Sequence | Product size | Temperature |
|---|---|---|---|
| GAPDH fwd | TCGGTGCGTGCCCAGTTGAAC | 64°C | 246 bp |
| GAPDH rev | ATGCGGCTGACTGTCGAACAGGAG | | |
| Sat 2 fwd | ATTCGAGTCCATTCGATGATTCCAT | 57°C | 398 bp |
| Sat 2 rev | ATGGAAATGAAAGGGGTCATCATCT | | |
| HoxA7 reg E fwd | CCCTCTATTCTCCATCGGAGAC | 60°C | 522 bp |
| HoxA7 reg E rev | CTGCACCATGTTGCACCAG | | |
| HoxA7 reg F fwd | CTGGTGCAACATGGTGCAG | 60°C | 497 bp |
| HoxA7 reg F rev | CCTCCTCCCGGACGCTG | | |
| Wp region 1 fwd | TTCATCATGTAACCCACAAATCA | 58°C | 405 bp |
| Wp region 1 rev | CGGAAGTGACACCAAATATCTCT | | |
| Wp region 2 fwd | CACTTCCGCATTTTAAGTTTCAG | 58°C | 431 bp |
| Wp region 2 rev | TCTGCTCGTTACCAGAGAGAATG | | |
| Wp reg 3 fwd | TCTGGTAACGAGCAGAGAAGAAG | 58°C | 572 bp |
| Wp reg 3 rev | ACAGAGAGAGGGGCAGAACC | | |
| Wp reg 4 fwd | TCTCTCTGTCCTTCAGAGGAACC | 58°C | 461 bp |
| Wp reg 4 rev | GTCTAGGGTGGAGCGAAGGT | | |
| Wp reg 5 fwd | GCTTCAGAGCCCAGGATGTC | 58°C | 473 bp |
| Wp reg 5 rev | GAGGCTGGACTTTACAGACAGTG | | |
| Wp reg 6 fwd | GTCTCCTGTGCACTGTCTGTAAA | 58°C | 484 bp |
| Wp reg 6 rev | AGATTTCGGGTCCAAATCACTAC | | |
| Wp reg 7 fwd | GACCCGAAATCTGACACTTTAGA | 58°C | 423 bp |
| Wp reg 7 rev | GAGGTAGAAGACCCCCTCTTACA | | |
| Wp reg 8 fwd | TGTAAGAGGGGGTCTTCTACCTC | 58°C | 469 bp |
| Wp reg 8 rev | ACTAAGCCTCCCTTTATGTGAGC | | |
| Wp reg 9 fwd | GCTCACATAAAGGGAGGCTTAGT | 58°C | 493 bp |
| Wp reg 9 rev | TCTGCTCGTTACCAGAGAGAATG | | |
| Cp fwd | AAATGTTGAGGGACCTAAGAGATG | 58°C | 414 bp |
| Cp rev | TGGCTTTAATTGTCATGTATGCTT | | |

**Table 2.5: List of primers used for X-ChIP**

### 2.13.7 Normalizing ChIP Q-PCR data

Two methods are commonly used to normalize ChIP Q-PCR data – the percent input method and the fold enrichment method. Analyzing the data relative to input is a better method as it included normalization for both background levels and input chromatin going into ChIP. An example of how each method is used to calculate the ChIP results is shown below.

### 2.13.7.1 Percent Input Method

In this method signals obtained from the ChIP are divided by those signals obtained from an input sample. The input sample represents the amount of chromatin used in the ChIP. Typically 1-10% of starting chromatin is used as input. An example of how this is calculated is shown below. Step1: As the starting input fraction is 10%, then a dilution factor of 10 0r 3.32 cycles (i.e., log2 of 10) is subtracted from the Ct value of the diluted input. Step 2; The percent input is calculated using the formula 100*2^(Adjusted input – Ct (IP). The % input is then plotted on the x axis against each antibody on the y axis.

| | Raw Ct | Ct adjusted to 100%<br>(Ct Input-3.32) |
|---|---|---|
| Input 10% | 25.917 | 22.59 |

Step 1

| | Average Ct | 100*2^(Adjusted input – Ct (IP) |
|---|---|---|
| Adjusted Input | 22.59 | |
| No Antibody | 28.864 | 1.3 |
| IgG | 28.625 | 1.53 |
| DNMT3A | 27.413 | 3.53 |

Step 2

### 2.13.7.2 Fold Enrichment Method

This normalization method is also called 'signal over background' or 'relative to the no antibody'. With this method the ChIP signals are divided by IgG antibody signals, representing the ChIP signal as the fold increase in signal relative to the background signal (as shown below). The assumption of this method is that the level of background signal is reproducible between different primer sets, samples and replicate experiments.

| | **Average Ct** | **Ct (IP) – Ct IgG** | **Ct (IP) – Ct IgG** |
|---|---|---|---|
| **IgG** | 28.625 | 0 | 0 |
| **No Antibody** | 28.864 | 0.239 | .85 |
| **DNMT3A** | 27.413 | -1.451 | 2.73 |

## 2.14 Immunohistochemistry

### 2.14.1 Preparation of cultured cells for immunohistochemistry

Adhesive-coated slides were assembled with Cytofunnel® disposable sample chambers, filter cards and Cytoclips™ (Thermo Electron) according to the manufacturer's instructions and inserted into the Shandon Cytospin centrifuge (Thermo Electron). 2 x $10^6$ cells were harvested and re-suspended in 1 ml PBS and fixed in 100 µl formalin. 100 µl of fixed cells were transferred into the Cytofunnels®. Cells were adhered to slides by centrifugation in the Cytospin at 1000 *g* for 5 minutes. Cells were air dried, fixed for 10 minutes in 10% formal-saline solution (Genta Medical), and air dried. Cytospin preparations were stored in foil at -20$^o$ C. When required, slides were thawed and washed in running tap water for 5 minutes before continuing with standard immunohistochemistry technique.

### 2.14.2 Preparation of tissue biopsy sections for immunohistochemistry

Sections of paraffin-embedded tissue biopsies were dewaxed and rehydrated using xylene and ethanol, 2 x 5 minutes each.

### 2.14.3 Blocking of endogenous peroxidase activity and antigen retrieval

Endogenous peroxidase activity was blocked for 10 minutes in 3% hydrogen peroxide in methanol (both Sigma-Aldrich). Slides were rinsed thoroughly in running tap water. Antigen retrieval was performed using the agitated low temperature epitope retrieval (ALTER) method by incubation of slides in EDTA buffer pH8 for 16 hours, and then cooled in running cold water for 5 minutes.

### 2.14.4 Detection of antigen

Slides were incubated in 1X PBS for 5 minutes, then blocked using 2X casein blocking solution (Vector laboratories), washed for 5 minutes in 1X PBS and incubated with diluted primary antibody for 1 hour. Samples were washed in PBS for 5 minutes. Secondary detection was with DakoChemate envision secondary antibody (Dako) using 2 drops of biotinylated universal secondary antibody was added to each tissue section for 30 minutes. Sections were washed again in PBS for 5 minutes.

### 2.14.5 Visualization and counterstaining

Visualization was carried out using diaminobenzidine (DAB) (Vector laboratories). A minimum of 100 μl of substrate solution was applied to each tissue section for between 30 seconds to 1 minute, during which time the substrate was converted to an insoluble brown product by the antigen-bound peroxidases. Slides were rinsed with PBS, counterstained with Mayer's haematoxylin (Sigma) for 10 seconds, washed under running tap water for 5 minutes, and dehydrated through two lots of ethanol and xylene for 5 minutes each, before being mounted with coverslips and one drop of DPX mounting (Invitrogen).

Chapter 3

# A genome wide analysis of transcriptional and methylation changes which follow Epstein Barr Virus infection of Germinal Centre B cells

**Aims of Chapter 3:**

To describe the transcriptional and DNA methylation changes associated with EBV infection.

**Objectives:**

1) To profile virus gene expression in lymphoblastoid cell lines (LCLs) established from GC B cells isolated from tonsillar tissue.
2) To describe the genome wide transcriptional changes which follow the infection of GC B cells with EBV and the extent to which these changes overlap with;
    i) those observed following the transfection of GC B cells with LMP1 and LMP2A
    ii) the transcriptional programme of transformed HL cell lines and of micro-dissected HR-S cells.
3) To describe the genome wide changes in DNA promoter methylation which follow infection of GC B cells with EBV.
4) To determine;
    i) if the distribution of methylation changes are related to chromosomal location, CpG content and baseline gene intensity scores.
    ii) the association between methylation and transcriptional changes.
5) The ontological associations of those genes found to be more or less methylated following EBV infection and their relevance to the pathogenesis of HL.

In this section, I first describe the establishment of lymphoblastoid cell lines from germinal centre B cells before profiling the expression of viral genes in these cells. I next describe the preparation of RNA, the assessment of its integrity and its subsequent hybridization to the transcriptional array platforms. After summarising the transcriptional changes observed in this array, I measure the extent to which the transcriptional profile of EBV infected GC B cells overlaps that of GC B cells transfected with LMP1 or LMP2A; that of HL cell lines when compared with GC B cells; and that of micro-dissected HRS cells compared with centrocytes and centroblasts.

## 3.1 Establishment of lymphoblastoid cell lines from germinal centre B cells and genome wide profiling of EBV associated transcriptional changes in these cells

One of the difficulties faced when studying the contribution of EBV to the early stages of its associated malignancies, is that the cell lines used to model disease are already fully transformed and rarely resemble their progenitor cell. Therefore, a cell line model which allows us to study the early consequences of EBV infection and their possible contribution to the pathogenesis of disease is needed. My disease focus is on HL, and therefore I chose to infect germinal centre B cells, the presumed progenitor cells of HL with EBV. This model was used to identify the global transcriptional and methylation changes that occur following infection and transformation.

### 3.1.1 Infection of freshly isolated GC B cells with 2089 virus particles

***Isolation of CD10$^+$ cells from tonsillar tissue:*** GC B cells were isolated from tonsillar tissue using CD10$^+$ selection and magnetic separation as described in Section 2.7. Following magnetic separation, more than 90% of the GC B cells were found to be CD10$^+$, comprising centroblasts (CD77$^+$) and centrocytes (CD77$^-$). Figure 3.1 shows the flow cytometric analysis of tonsillar mononuclear cells from one patient before and after enrichment.

***Infection of GC B cells:*** GC B cells isolated from three different tonsils were infected with EBV at a multiplicity of infection (MOI) of 50. A green fluorescent protein (GFP) tag on the virus was used to indicate the efficiency of viral infection which was found to be >80%, 48 hours post infection (Figure 3.2). Cells were harvested daily for the first four days following infection and then at weekly intervals up to 3 weeks with the last sample collected 6 weeks post infection. DNA, RNA and protein were prepared at each time point for subsequent analyses.

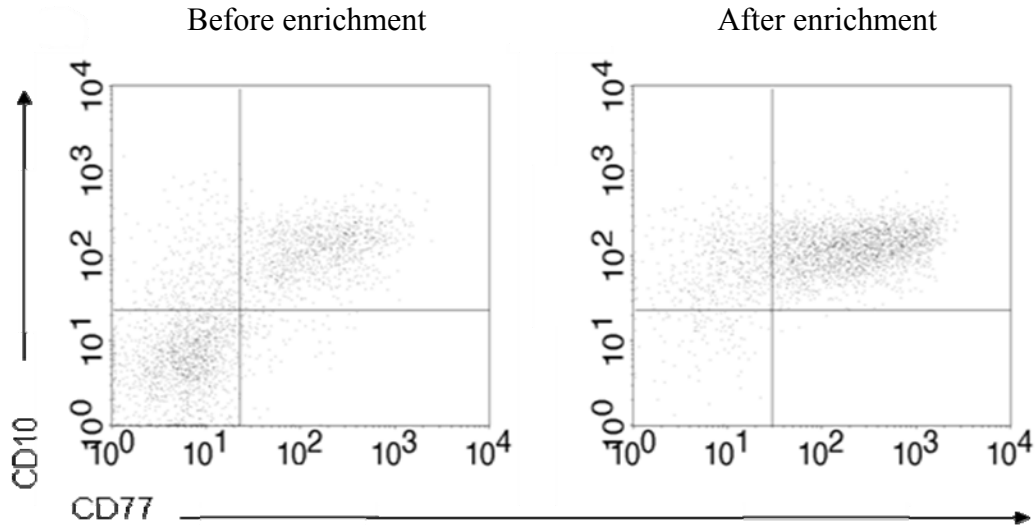
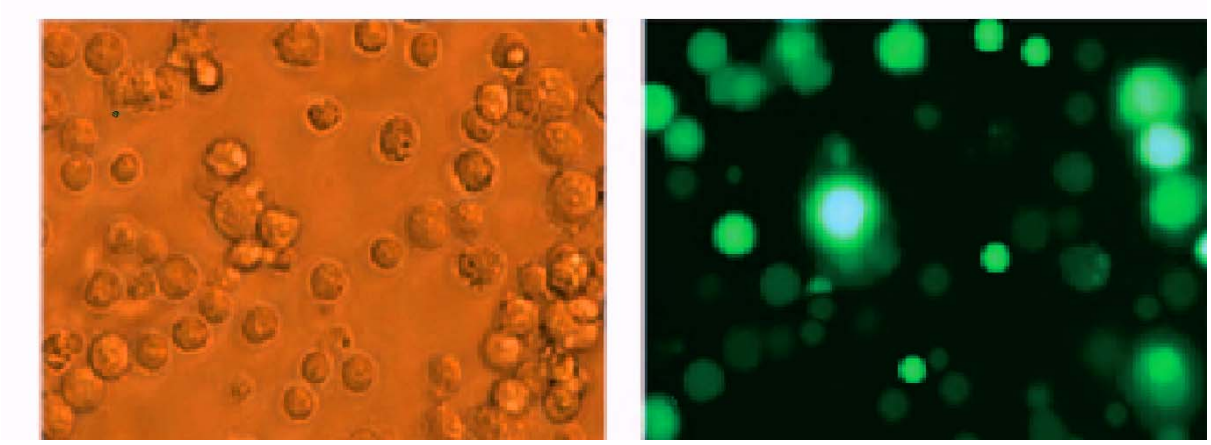



**Figure 3.1: Enrichment of CD10$^+$ GC B cells.** Flow cytometric analysis of tonsillar mononuclear cells before enrichment (left panel) and after (right panel), stained for CD10 and CD77.



Phase contrast 2089 expressing cells

**Figure 3.2: Infection of GC B cells with EBV.** Infection efficiency was determined using UV microscopy. Infected cells express GFP that can be detected upon exposure to UV light (right). Phase contrast (left) reveals the number of cells within the same area.

### 3.1.2 Gene expression profiling revealed a typical latency III pattern in infected GC B cells

Total RNA was isolated and reverse-transcribed into cDNA as described in Sections 2.2.1 and 2.4.2. Q RT-PCR was carried out using primers and probes specific for latent gene transcripts. The EBV transcript was normalised to the endogenous control, GAPDH, and results plotted relative to GC B cells. Although Figure 3.3 illustrates using data from three independent experiments, variation between LCLs established from different donors, the trends in viral gene expression are similar and can be summarised as follows.

- Wp is activated within 24 hours of infection, peaking between 48-72 hours, declining thereafter but remaining detectable at six weeks.
- Cp activation was first detected in one tonsil within 24 hours of infection and in the remaining two tonsils after 48 hours, increasing thereafter over the following six weeks.
- EBNA1 ($Y_3$-U-K) and EBNA2 transcripts, both indicators of Wp or Cp activity, were detectable 24 hours after infection, increasing in expression thereafter.
- There was no evidence of Q-U-K spliced EBNA1 mRNA indicative of Qp usage.
- LMP1 expression was detectable in all three tonsils within 24 hours of infection, increasing gradually thereafter.
- LMP2A, the last viral gene to be expressed in all three LCLs, was detected 72 hours post-infection.
- BZLF1 expression was low and variable across the three tonsils, suggesting that different numbers of cells were entering the lytic cycle.

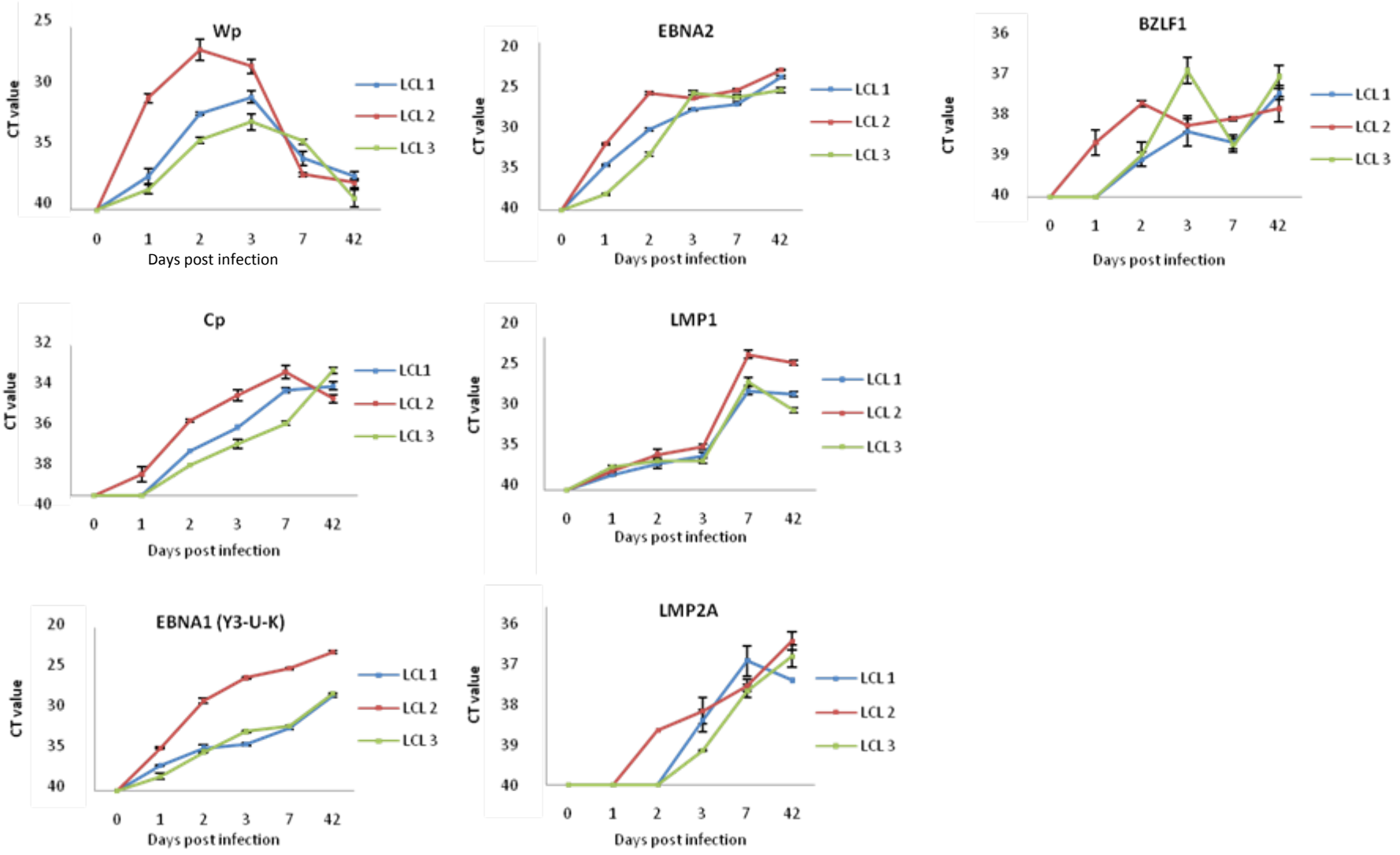



**Figure 3.3: Analysis of EBV gene expression following infection of GC B cells with EBV.** Q-RT PCR showing Wp, Cp, Y3-U-K-spliced EBNA1, EBNA2, LMP1, LMP2A and BZLF1 expression in three GC B cell derived LCLs at different time-points post infection. Assays were performed in triplicate and results are presented as raw Ct values.

### 3.1.3 Preparation of RNA for Transcriptional arrays

Total RNA from GC B cells (10 μg) and from LCLs (10 μg) six weeks post-infection with EBV was prepared for hybridization to Affymetrix GeneChip®Human Genome U133 Plus 2.0 arrays as described in Section 2.8. The RNA was used to synthesize double stranded cDNA which was then transcribed to cRNA. This cRNA was then fragmented, labelled and hybridized to the arrays. The efficiency of fragmentation was checked by running the product on a 1.5% agarose gel (Figure 3.4).

### 3.1.4 Assessment of RNA quality

Although the pre-processing RNA quality checks had been entirely satisfactory for both the uninfected GC B cells and their EBV infected counterparts, the number of genes recorded as present on these uninfected GC B cell arrays was substantially less than that found to be present on other GC B arrays performed within this Institute (Table 3.1). This was possibly due to an insufficient amount of good quality RNA prior to amplification of the GC B cell sample. Therefore I decided to discard this data set, and instead compared the transcriptional profile of EBV infected GC B cells with another set of unamplified GC B cells which had been used in an earlier array experiment performed in this institute (Vockerodt *et al.*, 2008).

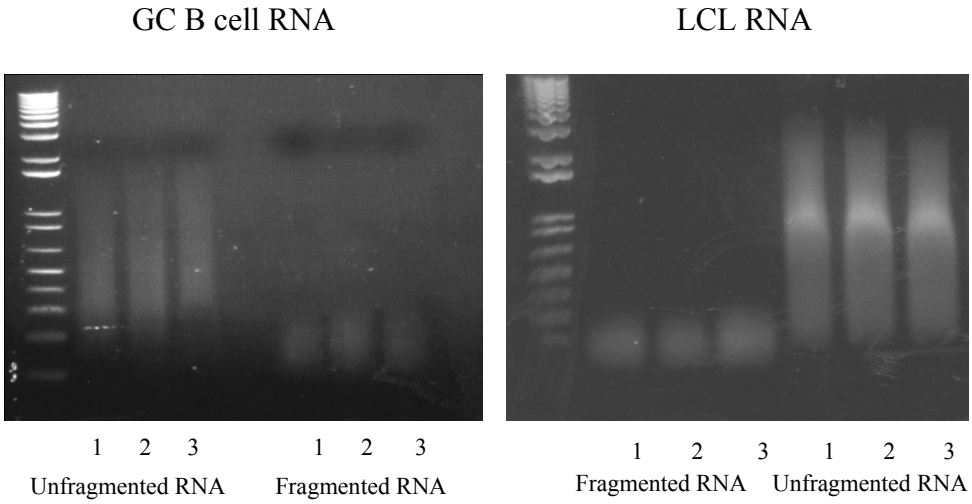



**Figure 3.4: Validation of RNA samples on an agarose gel**. An aliquot of fragmented and unfragmented RNA was run on an agarose gel to ensure the fragmentation step has worked successfully.

| | **Present on all 6 arrays** | **Absent on all 6 arrays** | **Absent on all 3 GC B cell arrays but present on all 3 LCL arrays** | **Absent on all 3 LCL arrays but present on all 3 GC B cell arrays** |
|---|---|---|---|---|
| **Matching GC B cell arrays** | 5562 | 29206 | 60 | 73 |
| **Alternative GC B cell arrays** | 15222 | 24184 | 1452 | 715 |

**Table 3.1: Summary of present or absent call on each array.** Probes that bind target transcripts above background noise are called present. Probes that do not bind target transcripts are called absent.

### 3.1.5 Measurement of gene expression changes in GC B cells following infection with EBV

Affymetrix GeneChip®Human Genome U133 Plus 2.0 arrays were used to analyze the gene expression changes following infection of GC B cells with EBV. These arrays allow for the analysis of 20765 genes with recognised gene symbols. The analysis of these arrays was performed with the help of Dr. Wenbin Wei, Head of the Bioinformatics department in the School of Cancer Sciences. The microarrays were scanned and the images analysed using Affymetrix microarray Suite 5.0. Probe level quantile normalization and robust multiarray analysis on the raw .CEL files were performed using the Affymetrix package of the Bioconductor (http://www.bioconductor.org) project.

### 3.1.6 Overall summary of transcriptional changes

Differentially expressed genes were identified using Significance Analysis of Microarrays (SAM). Using a false discovery rate (FDR) threshold of 5% with no fold - change cut off, a change in transcript binding was detected in 25474 probe sets; in 13548, there was an increase in binding and in 11926 a decrease. These probe sets mapped onto 13962 named genes; 6766 were up-regulated and 7196 were down-regulated (Table 3.2).

| Probe sets N = 25474 | | Named genes N = 13962 | |
|---|---|---|---|
| increased | decreased | increased | decreased |
| 11926 | 13548 | 6766 | 7196 |

**Table 3.2: Results of the SAM analysis performed using an FDR threshold of 5% on EBV infected GC B cells.**

### 3.1.7 Comparison of in-house and published arrays with EBV infected GC B cell arrays

The results of the EBV infected GC B cell arrays were compared with a number of in house and published arrays. In two of these arrays, GC B cells had been transfected with EBV latent genes, LMP1 or LMP2A (Vockerodt *et al.*, 2008, Vockerodt *et al.*, unpublished data). In the third, the transcriptional profile of four HL cell lines (L428, L591, KMH2 and L1236) had been compared with that of the GC B cells. I also compared the results of my array with a published dataset which compared the transcriptional profile of micro-dissected HRS cells with that of centrocytes and centroblasts (Brune *et al.*, 2008). All datasets contributing to these comparisons were re-analysed using SAM (5% FDR with no fold change cut-off), with the exception of the LMP2A array which was analysed using the Limma programme (5% FDR and 1.3 fold change cut off).

### 3.1.8 Assessment of significance of over-laps between transcriptional arrays

When comparing each of these data sets with my EBV infected GC B cell array, I first identified for each set of arrays, those genes that were concordantly up-regulated and those which were concordantly down-regulated. In order to exclude the possibility that the observed over-laps could have arisen by chance, I estimated the number of genes which I would have expected to be differentially expressed in both arrays given the frequency with which they were found to be differentially expressed in EBV infected GC B cells (Table 3.3). For example, of the 20765 genes on the U133 array, 6766 were up-regulated by EBV in GC B cells and 849 were up-regulated by LMP1 in GC B cells. By chance alone we would expect that some of the genes up-regulated by EBV in GC B cells would also be up-regulated by LMP1 in GC B cells. It might be expected that of the 849 genes up-regulated on the LMP1 GC B cell array, a proportion, 0.325 (6766/20765) would be up-regulated by EBV in GC B cells. However, rather than the 276 genes (0.325*849) that I would have expected to be up-regulated by both LMP1 and EBV in GC B cells, I found that in fact, 556 genes were up-regulated on both arrays. The difference between the observed and expected number of over-lapping genes provides the basis for significance testing using a simple chi-square test.

As well as measuring the significance of the association, we can also measure the magnitude of the association. A common statistic used to describe the magnitude of the association is the odds ratio. A worked example shown below compares the relative odds of being up-regulated by LMP1 in two groups of genes, one of which comprises genes up-regulated by EBV in GC B cells, the other genes which were transcriptionally unchanged following EBV infection. For those genes up-regulated by EBV, the odds of also being up-regulated by LMP1

were 0.091 or approximately 1 in 10 (556/6210). For those genes not up-regulated by EBV, the odds of being up-regulated by LMP1, the odds were 0.021 or approximately 1 in 50 (293/13706). The odds ratio is 4.3 (~0.091/0.021). Thus, there is greater than fourfold odds of a gene being up-regulated by LMP1 if it is also up-regulated by EBV than if it is not.

| worked example | | | |
|---|---|---|---|
| | | | |
| | | Up-regulated by LMP1 in GC B cells | |
| | | Yes | No |
| Up-regulated by EBV in GC B cells | Yes | 556 (a) | 6210 (b) |
| | No | 293 (c) | 13706 (d) |

The formula for the odds ratio can be written as: $OR = \frac{a/b}{c/d}$

Chi-square tests were performed and odds ratios estimated for each of the array comparisons and the results are summarised in Table 3.3. As a control of convenience, an array performed on GC B cells transfected with BMI1, a component of the Polycomb group (PcG) multiprotein PRC1 complex, was included in Table 3.3. As might have been expected, we do not see a significant over-lap between genes differentially regulated by BMI1 and EBV in GC B cells. The most compelling result from Table 3.3 is the finding that genes which were differentially expressed in HL cell lines compared with GC B cells were substantially and significantly enriched for those that were differentially regulated in GC B cells infected with EBV. A 60% over-lap in concordantly regulated genes was observed between the micro-dissected HRS cells and the EBV infected GC B cells. These results indicate that EBV infected GC B cells may be a good model for recapitulating those early changes associated

with HL. Not surprisingly the over-lap between the EBV infected GC B cells and the LMP1 transfected GC B cells was substantial and significant given that LMP1 is the major transforming oncogene of EBV.

| Frequency with which genes found to be differentially expressed in GC B cells following infection with EBV were also found to be differentially expressed in HL cell lines, HRS cells and GC B cells transfected with LMP1 or LMP2A | | | | | | | |
|---|---|---|---|---|---|---|---|
| | **Array comparison** | **Increased on array** | **Increased on array and in EBV infected GC B cells** | **% concordantly increased** | **Odds Ratio** | **95% CI** | ***p* value** |
| **Increased in LCL n = 6766** | HL cell lines vs. GC B cells | 1724 | 1313 | 76% | 8 | 7.1 to 9 | p = 0.0000 |
| | HRS cells vs. centrocytes | 6569 | 2558 | 39% | 1.5 | 1.4 to 1.6 | p = 0.0000 |
| | HRS cells vs. centroblasts | 4304 | 1856 | 43% | 1.8 | 1.7 to 1.9 | p = 0.0000 |
| | LMP1-GC B vs. GC B cells | 849 | 556 | 65% | 4.2 | 3.6 to 4.8 | p = 0.0000 |
| | LMP2A-GC B vs. GC B cells | 310 | 156 | 50% | 2.1 | 1.7 to 2.6 | p = 0.0000 |
| | BMI1-GC B vs. GC B cells | 139 | 17 | 12% | 0.28 | 0.17 to 0.47 | p = 0.0000 |
| | **Array comparison** | **Decreased on array** | **Decreased on array and in EBV infected GC B cells** | **% concordantly decreased** | **Odds Ratio** | **95% CI** | ***p* value** |
| **Decreased in LCL n = 7196** | HL cell lines vs. GC B cells | 2207 | 1764 | 80% | 9.6 | 8.6 to 10.7 | p = 0.0000 |
| | HRS cells vs. centrocytes | 2292 | 1375 | 60% | 3.3 | 3 to 3.6 | p = 0.0000 |
| | HRS cells vs. centroblasts | 3607 | 2017 | 56% | 2.9 | 2.7 to 3.2 | p = 0.0000 |
| | LMP1-GC B vs. GC B cells | 1347 | 860 | 64% | 3.6 | 3.2 to 4.1 | p = 0.0000 |
| | LMP2A-GC B vs. GC B cells | 440 | 204 | 46% | 1.7 | 1.5 to 2.1 | p = 0.0000 |
| | BMI1-GC B vs. GC B cells | 229 | 73 | 31% | 0.88 | 0.66 to 1.1 | p = 0.3714 |

**Table 3.3: Frequency with which genes found to be differentially expressed in GC B cells following infection with EBV were also found to be differentially expressed in HL cell lines, HRS cells and GC B cells transfected with LMP1 or LMP2A:** Significance testing was carried out using a chi-square test (www.quantitativeskills.com/sisa/statistics/twoby2.htm). The BMI1-GC B vs. GC B cells analysis serves as a control, as we would not expect to see a significant number of genes concordantly regulated between GC B cells transfected with BMI1 and GC B cells infected with EBV.

**Summary**

In this section, I have described how I established three LCLs from different patient GC B cells. I have confirmed that these LCLs express the typical Latency III pattern of viral gene expression. I have also shown that the transcriptional profile of EBV infected GC B cells overlaps that of HL.

## 3.2 Methylation arrays

In this section, I describe how I first isolated and then hybridized to a promoter array, methylated DNA from EBV infected GC B cells. Having provided a summary of the genome wide changes in the methylation status of cellular genes, I next confirmed some of these changes using pyrosequencing and a candidate gene approach. Finally, I determine for selected genes, whether the methylation changes observed at six weeks are progressive on further cultivation of EBV infected GC B cells.

### 3.2.1 Optimization of the immunoprecipitation assay

Whole genomic DNA was isolated from both the LCLs and GC B cells as described in Section 2.2.3. Methylated DNA was isolated using an immunoprecipitation method (MeDIP) originally described by Weber *et al,*. 2005. For this MeDIP assay, a monoclonal 5-methyl-cytosine antibody was used to immunocapture methylated genomic fragments.

***Sonication***: DNA was sonicated and conditions optimised so as to yield fragments between 200 - 1500 bp in length. The optimum conditions were provided by 20 minutes of 30 second pulses at the highest setting, with 30 seconds on ice in between each pulse (Figure 3.5).

***Determining the DNA-to-antibody ratio:*** Varying amounts of antibody (1 µg to 15 µg) were added to 5 µg of LCL DNA and the methylated DNA recovered from the elution fraction as described in Section 2.9. The relative enrichment in the eluant was compared to input fractions was calculated using SYBR green Q PCR as described in Section 2.9.4. Enrichment in each MeDIP assay was determined using primers specific for the methylated Xist sequence, (Figure 3.6). I concluded that 5 µg of input genomic DNA incubated with 7.5 µg of 5-

methyl-cytosine antibody achieved optimal recovery of methylated DNA in the LCLs. All subsequent MeDIP experiments were performed with this DNA-to-antibody ratio.

### 3.2.2 Immunoprecipitation of LCL and GC B cell DNA

To allow for the variation that may occur in a single immunoprecipitation, MeDIPs were carried out in triplicate for each of the three GC B cell and LCL samples. Methylated DNA was assayed in triplicate using SYBR green Q-PCR as previously described in section 2.9.4. Enrichment in each elution fraction was determined using primers specific for the methylated Xist sequence, (this enrichment varied from approximately 50 – 75 % across the six samples). In contrast, the unmethylated GAPDH sequence formed less than 5% of the elution fraction in all six samples (Figure 3.7). Following confirmation of enrichment for each MeDIP performed, the three elution fractions for each individual sample were pooled.

### 3.2.3 Amplification of immunoprecipitated DNA

Inevitably methylated DNA comprises only a fraction of the DNA in the starting sample. However, 7.5 μg of DNA is needed for hybridization to the promoter methylation arrays. Therefore, it was necessary to amplify the methylated DNA. Methylated DNA isolated from each of the three patient GC B cells and LCLs along with input DNA that did not undergo MeDIP was amplified in triplicate using round A/B/C random amplification. (Although, alternative whole genome amplification methods were considered and tried at this point, the A/B/C random amplification method was found to be the most successful and is described in full in Section 2.9.5). To ensure that enriched methylated DNA had not been lost during amplification; amplified products were again checked for enrichment of methylated DNA.

Amplified products from each sample were then pooled, and further processed in preparation for hybridisation to the arrays as described in Section 2.9.6.

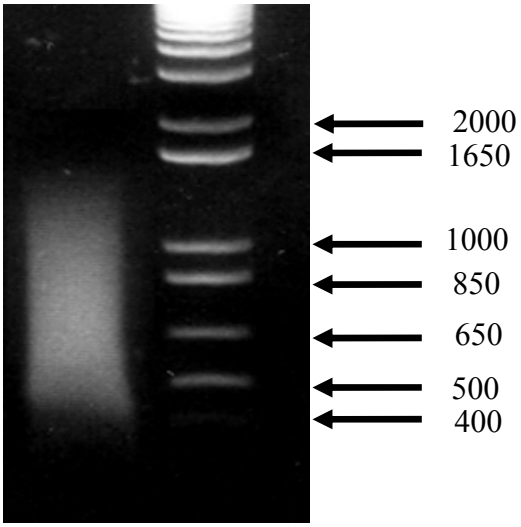



**Figure 3.5: Sonication of LCL DNA:** Sonication conditions were optimised to obtain DNA fragments of 700 bp in average length.

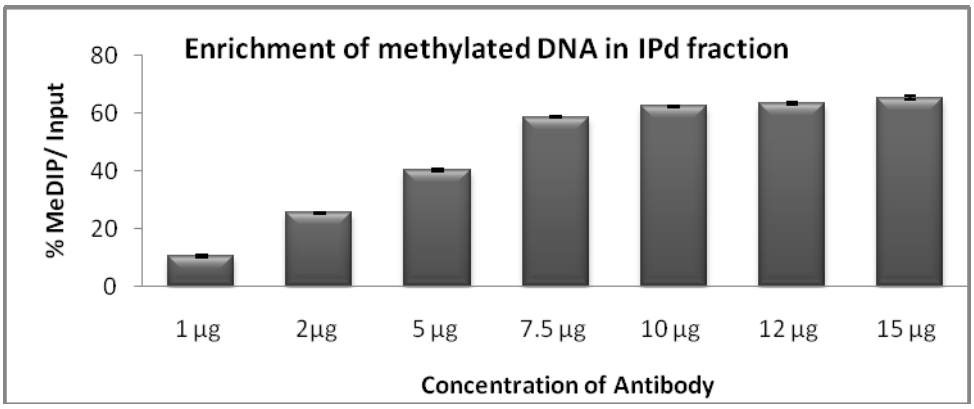



**Figure 3.6: Optimal DNA to antibody ratio:** The optimal amount of antibody required to pull out methylated sequences from 5 µg DNA was calculated by varying the amount of antibody and performing MeDIP on each sample. The eluted fractions were assayed using SYBR green Q PCR with the methylated XIST sequence. Each reaction was performed in

triplicate and the percentage MeDIP/Input was calculated as described in Section 2.9.4 of Materials and methods.

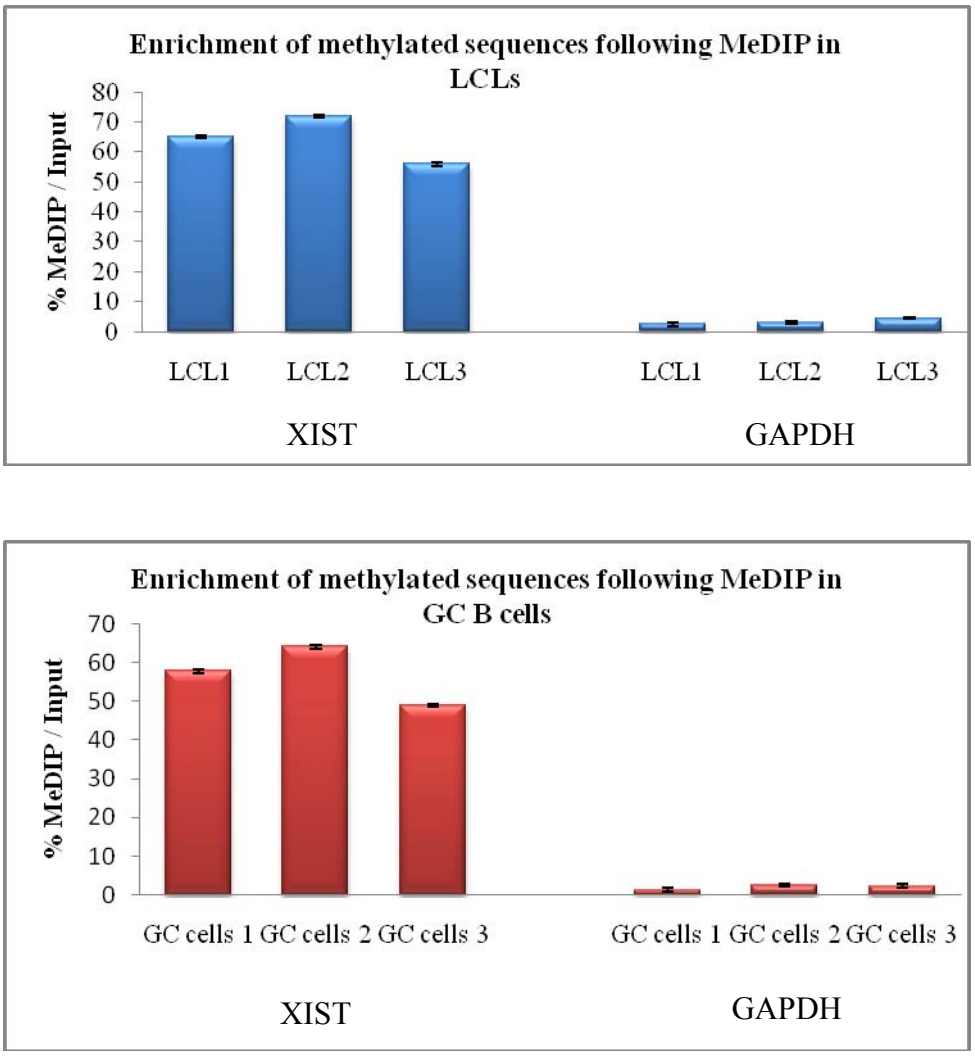



**Figure 3.7: Enrichment of methylated and unmethylated sequences following MeDIP of LCL and GC B cell samples.** Enrichment of methylated DNA was assayed using SYBR green Q PCR. GAPDH served as an unmethylated control and Xist served as a methylated control. Each Q PCR reaction was carried out in triplicate. % MeDIP / Input was calculated as described in Section 2.9.4 of Materials and Methods.

### 3.2.4 Methylation array analysis

Following hybridization and scanning of the enriched GC B cell, LCL samples and input DNA, the raw array data from each of the biological replicates were quantile-normalized. Following normalization, a software programme, TileMap, was used to identify regions of hypermethylation and hypomethylation.

#### 3.2.4.1 Setting parameters for analysis using TileMap software

The normalised data were analysed using TileMap, a software package designed for the analysis of tiling arrays. This package offers statistical models and algorithms for the detection of genomic loci showing hybridization patterns of interest. It uses a test-statistic that is computed for each probe based on a hierarchical empirical Bayes model. Combination of test-statistic information from neighbouring probes within a genomic region can be used to show whether the region has a hybridization pattern of interest (Ji and Wong, 2005). Two different methods, the moving average (MA) method and the hidden Markov model (HMM) method, were used to analyze normalized data from the methylation arrays. For the MA method, window size was set to 11 probes and maximum gap allowed between each set of 11 probes, set to 300bp. For the HMM method, expected hybridization length was set to 11 probes, and the maximal gap allowed set to 200bp. For both methods, peaks were merged if the gap between two peaks were less than the maximal gap allowed and the number of probes that failed to pass cut off between the two peaks less than 6. Peaks were discarded if peak length was less than 100bp or did not contain at least 5 continuous probes passing the cut off. When calculating the false discovery rate (FDR) for the HMM method the unbalanced mixture subtraction (UMS) was used; “left tail” was used for the MA method. All probe sequences were mapped to the genome assembly (Hg18). The output from TileMap included

final summaries for each probe and a *.bed file containing selected genomic regions that could be visualized using the Affymetrix Integrated Human Genome browser (IGB).

#### 3.2.4.2 Analysis of methylation changes in EBV infected GC B cells

These biological samples allowed for a number of different comparisons. For example, a comparison of the GC B cell arrays with the input DNA arrays allowed the identification of genes that are hypermethylated or hypomethylated in GC B cells. However, because I wanted to identify genes that had a change in methylation status following infection of GC B cells with EBV, my primary interest was in the direct comparison of the GC B cell arrays with LCL arrays. Therefore, this is the only analysis presented in the main body of my thesis.

### 3.2.5 Promoter methylation arrays predict widespread changes following EBV infection of GC B cells

The number of genes with a change in methylation status varied according to the analytical method used. Methylation changes were detected in 1783 genes using the MA method of analysis, and 506 genes were identified using HMM (Figure 3.8). Of the 506 genes with methylation changes identified using HMM, 83% were found to be concordantly methylated using the MA method (Table 3.4). I decided that my preferred method of analysis would be predicated on the results of the validation experiments described below.

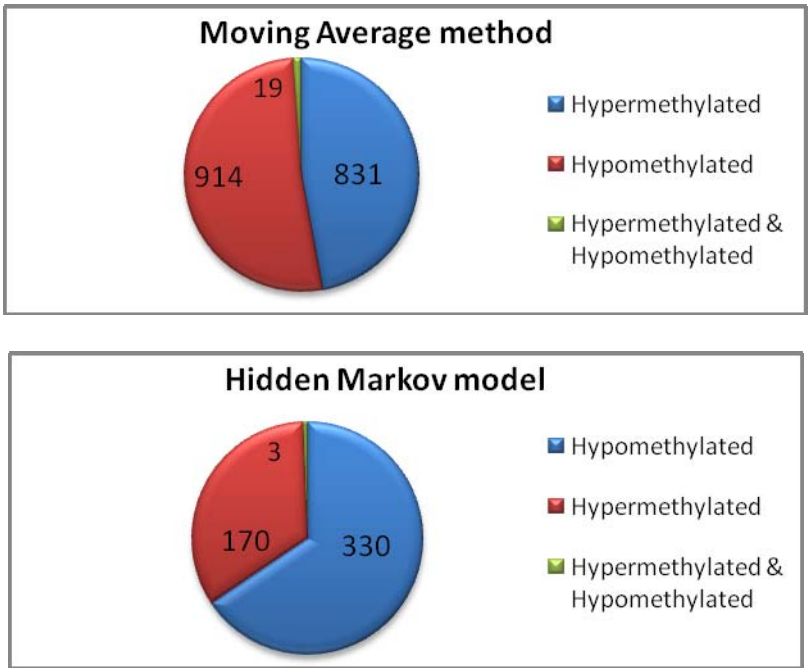



**Figure 3.8: Number of genes with a change in methylation status upon EBV infection of GC B cells**. Identified hypermethylated genes from each analysis are shown in red and hypomethylated genes reported in blue. Genes reported as containing sites of both hypermethylation and hypomethylation are reported in green.

| | | **Hidden Markov model** | | | | |
|---|---|---|---|---|---|---|
| | Change in methylation status | LCL > GCB | LCL < GCB | LCL > and < GCB | LCL = GCB | **Total** |
| **Moving average** | LCL > GCB | 106 | 0 | 0 | 725 | **831** |
| | LCL < GCB | 1 | 307 | 0 | 606 | **914** |
| | LCL > and < GCB | 2 | 3 | 3 | 11 | **19** |
| | LCL = GCB | 61 | 20 | 0 | 18105 | **18186** |
| | **Total** | **170** | **330** | **3** | **19447** | ***19950*** |

**Table 3.4: Correspondence between HMM and MA methods of analysis.** The number of genes reported as hypermethylated or hypomethylated, or both using MA and HMM methods of analysis. The total number of genes on the promoter methylation array is 19950.

### 3.2.6 Validation of methylation changes predicted on the array

Pyrosequencing, a quantitative method for analysing methylation at individual CpGs, was used to validate the results of the methylation array. 12 genes were chosen for validation; in 8 of these, methylation change was predicted by both the HMM and MA method, in 2 by the MA alone and in 2 by the HMM alone.

#### 3.2.6.1 Criteria for selecting candidate genes for validation

In order to meet some of the other objectives of this thesis one or more of the following criteria were used to select candidate genes for further validation:

1. transcriptional change on the expression array
2. differentially expressed in HL compared with GC B cells
3. differentially expressed in GC B cells following transfection with LMP1 or LMP2A.
4. methylation known to be increased or decreased in malignancy

These criteria are set out in detail for each of the candidate genes in Table 3.5

#### 3.2.6.2 Pyrosequencing confirms array predictions

Before designing primers for pyrosequencing, the integrated genome browser (IGB) was used to visualize regions predicted as hypermethylated or hypomethylated on the array. The UCSC genome browser was then used to obtain the DNA sequences for the genes of interest. Pyrosequencing primers and assays were designed as described in Section 2.10.2. To confirm EBV-associated changes in methylation status, pyrosequencing was performed on each of the

three GC B cell and LCL samples. Illustrative examples of pyrosequencing results for selected candidate genes are now provided.

| Predicted methylation change on promoter array | Gene symbol | EBV infected GC B cell transcriptional array | Expression in HL cell lines or HRS cells | Regulated by LMP2A or LMP1 | Regulated by methylation | Associated to pathogenesis |
|---|---|---|---|---|---|---|
| **Hypomethylated** | CSMD1 | No change | Increased in HRS cells (Brune *et al* 2008) | Neither | Inactivated by methylation (Richter *et al* 2005) | TSG (Ma *et al* 2009) |
| | SPRY2 | Not present | Increased in some HL cell lines (Vockerodt *et al* 2008) | Neither | Decreased upon demethylation of L428 and KMH2 ( in house arrays). Methylated in B cells (Frank *et al* 2009) | Its expression inhibits tumour growth in diffuse large B-cell lymphoma (DLBCL) (Sanchez *et al* 2008) |
| | MAGEA3 | Not present | Increased in HL cell lines (Vockerodt *et al* 2008) | Neither | Increased following demethylation of L428 and KMH2 (in house arrays). Promoter hypomethylation in colorectal carcinomas and gastric cancers (Kim *et al* 2006) | Expressed in malignant tumours and precancerous lesions but not in benign tumours (Kim *et al* 2006) |
| | PRDM1 | Decreased | Decreased in HL cell lines Down-regulated in HRS cells (Nie *et al* 2008) | Decreased by LMP2A Decreased by LMP1 | No | May contribute to the phenotype maintenance and pathogenesis of HRS cells |
| | ICMT | Increased | Increased in HRS cells Increased in all HL cell lines | Neither | No | Inactivation of *Icmt* ameliorates phenotypes of K-RAS–induced malignancies in vivo (Wahlstrom *et al* 2008) |
| | FGFR2 | Increased and decreased | Increased in HRS cells Increased in 3 HL cell lines, decreased in 2 HL cell lines | Neither | Hypomethylated in gastric cancer (Nishigaki *et al* 2005) | Deregulation of the FGFR2 gene has been identified in a number of cancers (Nan *et al* 2009) |
| | GRB10 | Increased | Increased in HRS cells Increased in all HL cell lines | Neither | Methylation important for imprinting (Sanz *et al* 2008) | GRB10 is a protooncogene up-regulated in AML (Casas *et al* 2003) |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Hypermethylated** | ELL3 | Decreased | Decreased HRS cells (Brune *et al* 2008) | Decreased by LMP1 | No | ELL3 can regulate cell growth and survival and ELL translocations result in the development of human malignancies. (Johnstone *et al* 2001) |
| | TCL6 | Decreased | Decreased in HRS cells<br>Decreased in all HL cell lines | Decreased by LMP1 | No | The association of this gene with T-cell leukemia chromosome translocations implicates this gene as a candidate for leukemogenesis (Saitou *et al* 2000). |
| | RBM5 | Decreased | Decreased in HRS cells<br>Decreased in all HL cell lines | Decreased by LMP1 | No | RBM5 is a TSG and loss of expression may increase the metastatic potential of tumours (Oh *et al* 2010 |
| | SMAD4 | Decreased | Decreased in HRS cells<br>Decreased in some HL cell lines | Neither | Promoter methylation correlates with reduced SMAD4 expression in advanced prostate cancer (Aitchison *et al* 2008) | TSG (Liu et al 2001) |
| | ID2 | Increased | Increased in all HL cell lines | Increased by LMP1 | No | ID2 is a suppressor of B-Cell-Specific Gene Expression, in Hodgkin's Lymphoma (Renne *et al* 2006) |

**Table 3.5: Candidate genes for validation by pyrosequencing**

1) The array predicted hypomethylation of the cancer testis antigen, MAGEA3 following infection of GC B cells with EBV (Figure 3.9 panel A). In the same figure (panel B), a pyrogram shows for each of the CpGs examined that the % methylation was less in the LCL than that observed in GC B cells; the average fall in methylation across these 4 CpG sites was 9.75%.

2) The array predicted hypermethylation of the elongation factor RNA polymerase II-like 3 gene, ELL3, following infection of GC B cells with EBV (Figure 3.10 panel A). In the same figure (panel B), a pyrogram shows that for three of the four CpGs examined; the % methylation was more in the LCL than in GC B cells; the % increase was substantial in one CpG (highlighted) but only modest in the other two.

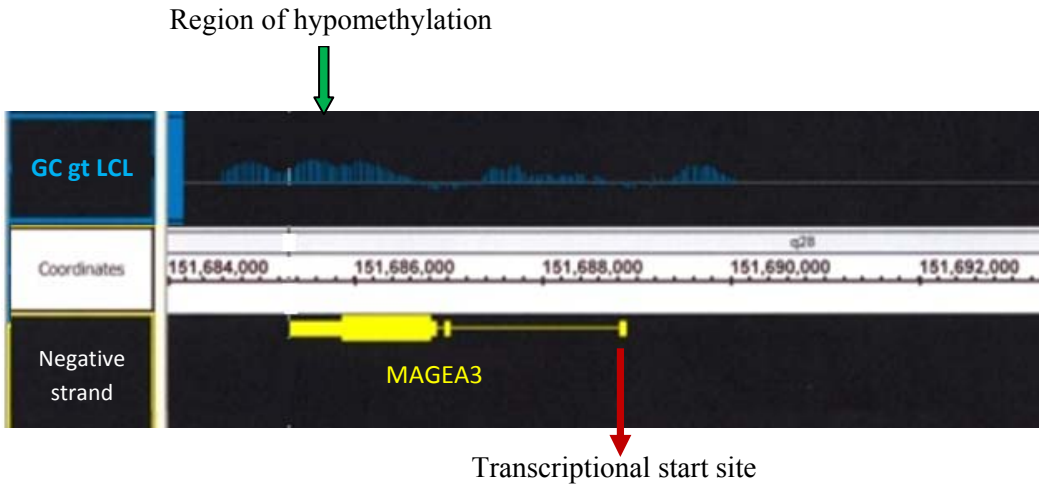
When summarising the pyrosequencing results for the candidate genes, I report both the average methylation status across all CpG sites and the maximum change observed at individual sites. Table 3.6 reports the pyrosequencing results for all the genes chosen as validation targets. The IGB ideograms and pyrograms for each gene are shown in Appendix 1. All seven genes predicted as hypomethylated on the array had their hypomethylation status confirmed by pyrosequencing. Four out of the five genes predicted as hypermethylated on the array were also found to be unambiguously hypermethylated using pyrosequencing analysis. All genes predicted to be hypermethylated or hypomethylated on the array using either analysis were confirmed using pyrosequencing. I chose to use the MA dataset for all further analyses because this method identified more genes with a change in methylation status compared to the HMM method.

| Hypomethylated genes | | CpG 1 | CpG 2 | CpG 3 | CpG 4 | CpG 5 | CpG 6 | CpG 7 | Average methylation |
|---|---|---|---|---|---|---|---|---|---|
| **CSMD1** | GC B cell | 98 | 89 | 87 | 83 | | | | **89** |
| | LCL | 94 | 80 | 70 | 63 | | | | **77** |
| **SPRY2** | GC B cell | 14 | 12 | 12 | 16 | 13 | 16 | | **13.8** |
| | LCL | 2 | 0 | 2 | 7 | 0 | 6 | | **2.8** |
| **MAGEA3** | GC B cell | 18 | 24 | 37 | 39 | | | | **29.5** |
| | LCL | 10 | 18 | 29 | 22 | | | | **19.75** |
| **PRDM1** | GC B cell | 31 | 29 | 30 | 24 | 26 | | | **28** |
| | LCL | 8 | 7 | 14 | 12 | 9 | | | **10** |
| **ICMT** | GC B cell | 44 | 49 | 48 | 88 | 54 | | | **56.6** |
| | LCL | 11 | 13 | 6 | 26 | 7 | | | **12.6** |
| **FGFR2** | GC B cell | 74 | 65 | 72 | 76 | 100 | 86 | 51 | **74.8** |
| | LCL | 23 | 62 | 64 | 56 | 94 | 78 | 50 | **61** |
| **GRB10** | GC B cell | 48 | 91 | 100 | 100 | 80 | 74 | | **82.2** |
| | LCL | 20 | 73 | 99 | 100 | 85 | 65 | | **73.6** |

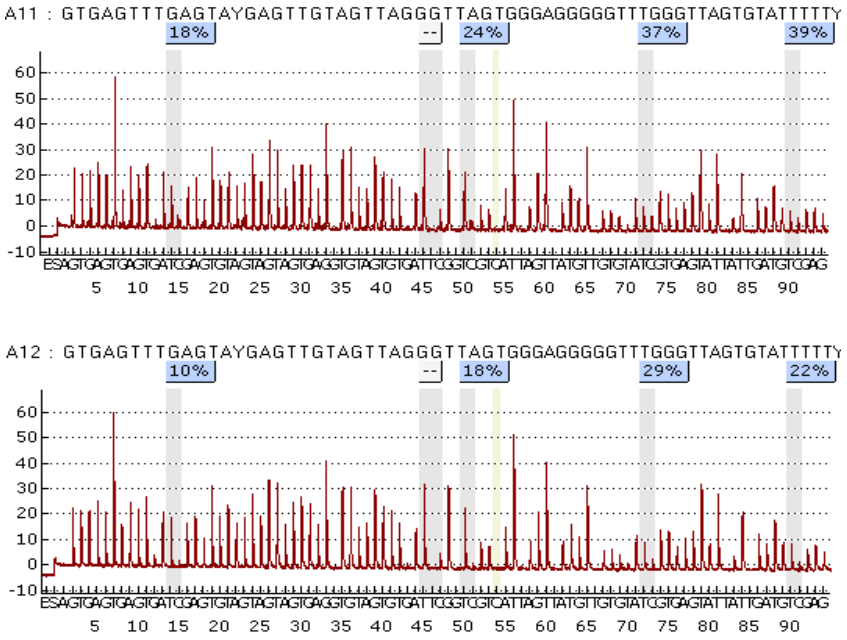
| Hypermethylated genes | | CpG 1 | CpG 2 | CpG 3 | CpG 4 | CpG 5 | CpG 6 | CpG 7 | Average methylation |
|---|---|---|---|---|---|---|---|---|---|
| **ELL3** | GC B cell | 97 | 62 | 100 | 73 | | | | **83** |
| | LCL | 97 | 94 | 100 | 75 | | | | **91.5** |
| **TCL6** | GC B cell | 31 | 68 | 56 | 100 | 21 | | | **55.2** |
| | LCL | 57 | 95 | 88 | 100 | 52 | | | **78.4** |
| **RBM5** | GC B cell | 4 | 0 | 4 | 6 | 4 | | | **3.6** |
| | LCL | 11 | 6 | 4 | 6 | 4 | | | **6.2** |
| **SMAD4** | GC B cell | 62 | 63 | 50 | 89 | 100 | | | **72.8** |
| | LCL | 75 | 70 | 57 | 92 | 100 | | | **78.8** |
| **ID2** | GC B cell | 69 | 69 | 64 | 69 | 61 | 31 | | **60.5** |
| | LCL | 56 | 90 | 87 | 56 | 52 | 26 | | **61.2** |

**Table 3.6: Summary of pyrosequencing results:** The percentage methylation at each CpG is recorded along with the average methylation across each region for both the GC B cells and the LCLs. The maximum difference highlighted in yellow. In each case, pyrosequencing was performed on three independently derived LCLs and their counterpart GC B cells.

## A. Array prediction for MAGEA3




## B. Pyrosequencing results

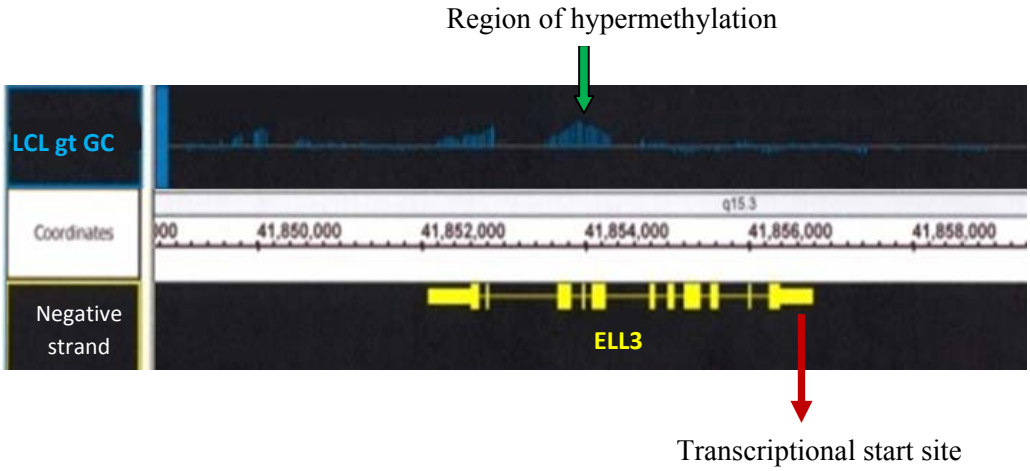



Average GC B cell methylation = 29.5 %

Average LCL methylation = 19.75 %

**Figure 3.9: IGB and pyrosequencing results for MAGEA3. A.** IGB results for the hypomethylated gene MAGEA3. The transcriptional start site is indicated by a red arrow and the region hypomethylated is indicated with a green arrow. **B.** Pyrosequencing results for the hypomethylated region indicated with the green arrow in IGB. GC B cells (top panel) and LCL (lower panel). The average methylation for the region analysed is shown on the right and the % methylation for each CpG is shown in the blue box above the CpG. Pyrosequencing was performed on three independently derived LCLs and their corresponding GC B cells.

**A.** **Array prediction for ELL3**




**B.** **Pyrosequencing results**

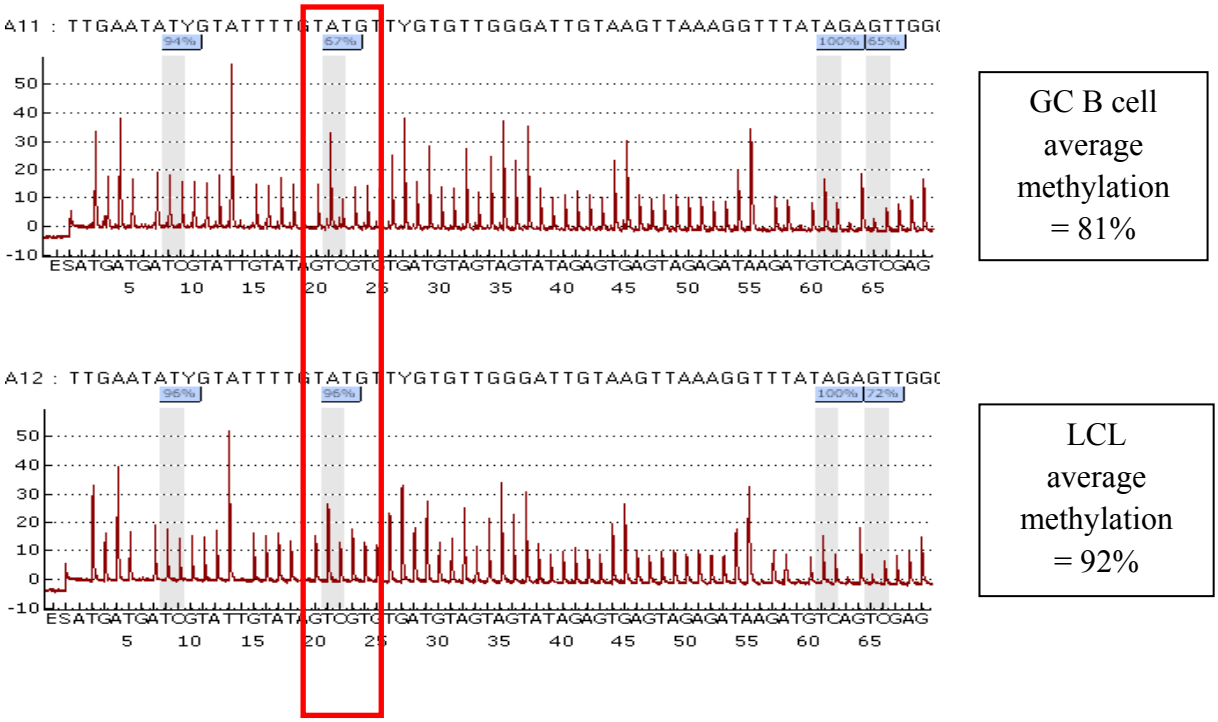



**Figure 3.10: IGB and pyrosequencing results for ELL3.** **A.** IGB results for the hypermethylated gene ELL3. The transcriptional start site is indicated by a red arrow and the region hypermethylated is indicated with a green arrow. **B.** Pyrosequencing results for the hypermethylated region indicated with the green arrow in IGB. GC B cells (top panel) and LCL (lower panel). The average methylation for the region analysed is shown on the right and the % methylation for each CpG is shown in the blue box above the CpG. Highlighted with the red box is the second CpG within the sequence being analysed. Pyrosequencing was performed on three independently derived LCLs and their corresponding GC B cells.

### 3.2.6.3 EBV induced methylation changes are progressive

Although the methylation arrays were carried out on LCLs six weeks post infection I was able to analyse methylation in the same LCLs kept in culture for a further six weeks. This provided the opportunity to explore whether methylation changes were progressive on further culturing. For example, serial pyrosequencing shows that the methylation of one of the CpGs in RBM5 increased from 4% at baseline to 11% at six weeks and to 20% at three months post infection (Figure 3.11). Similar changes in methylation were found in the other two genes included in this anlaysis, the results are reported in Annex 2.

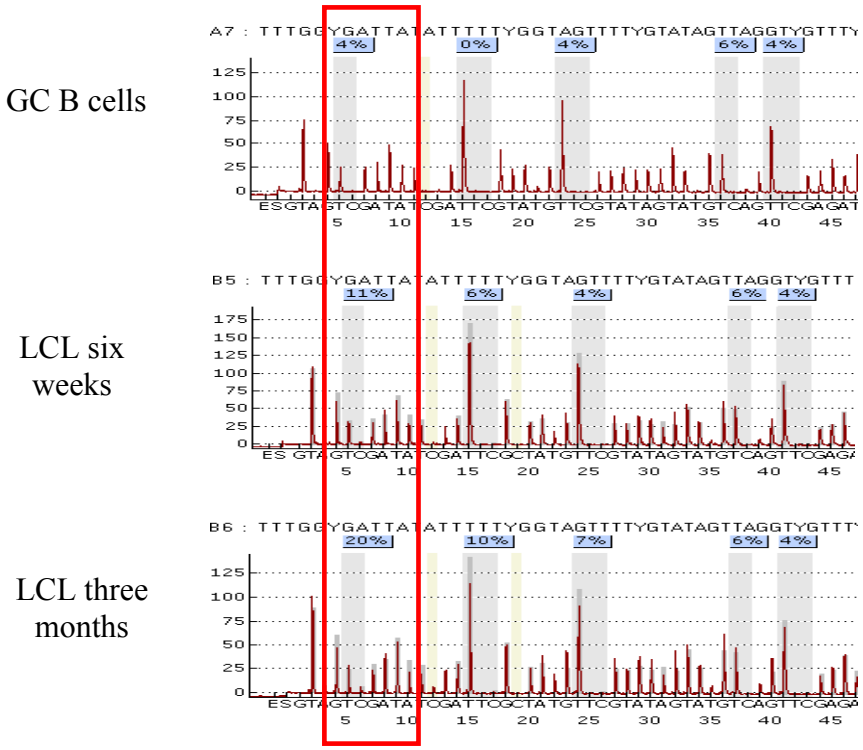



**Figure 3.11: Increased methylation over time.** Pyrosequencing of GC B cells (top panel), six week LCLs (middle panel) and three month LCLs (bottom panel). The first CpG analysed is highlighted with the red box and is highlighted because it has the greatest change.

**Summary**

In this section, I have described how I first isolated and then hybridized to a promoter array, methylated DNA from EBV infected GC B cells. I have confirmed using pyrosequencing, methylation changes predicted on the array, and shown that methylation changes are progressive on further long-term cultivation of EBV infected GC B cells.

## 3.3 Analysis of the distribution and determinants of EBV associated methylation events and their relationship to transcriptional change

In this section, I first examine whether the methylation changes following EBV infection of GC B cells are randomly distributed throughout the genome. I then investigate the determinants of these methylation changes. Specifically, I consider whether methylation changes can be related to CpG content or baseline gene expression. Finally, using the methylation and transcriptional arrays, I determine how often changes in methylation status predict changes in expression.

### 3.3.1 Methylation changes are commonly concordant in adjacent genes

I first investigated whether those genes with increased methylation following EBV infection were significantly more likely to be found immediately adjacent to at least one other hypermethylated gene. In order to do this, I first had to identify which genes listed on the Affymetrix promoter array were immediately adjacent to each other. I excluded 1596 genes listed on the array because their transcriptional start site fell between the transcription start and end site of the immediately preceding gene. All of the remaining genes (n = 18354) on the Affymetrix promoter array were then ordered according to their transcriptional start site and numbered one to 18354. These genes included 751 genes hypermethylated and 834 hypomethylated following infection with EBV. Of the 751 genes hypermethylated by EBV, 177 were found to be immediately adjacent to another hypermethylated gene (Table 3.7). In order to determine if this number of pairs was greater than what would have occurred by chance, I next measured how many concordantly methylated pairs would have been found

were I to select random samples of size equal to the number of hypermethylated (751) genes predicted on the array. Dr. Wenbin Wei provided me with statistical support for this aspect of the analysis. The methods employed are now outlined in brief using hypermethylated genes as an illustrative example. The computer programme R was used to generate 100000 random samples of size (n = 751) from the total population of 18354 genes listed on the array, and for each pair within each random sample, it was then determined whether the pair was consecutive. The distribution of the number of adjacent pairs in 100000 samples each containing 751 genes randomly selected from the total of 18354 genes is shown in Figure 3.12. I next calculated the probability of observing within a sample of 751 randomly selected genes, a number of consecutively numbered genes greater than 177 (the number of hypermethylated genes found to be adjacent to each other). Given that my simulation exercise suggested that the maximum number of consecutively numbered genes in a 100000 trials was only 58, I conclude that the probability of observing a 177 adjacent genes which were also hypermethylated is less than 1 in 100000 (Table.3.8)

| **Change in methylation status** | **Number of concordantly regulated pairs** |
|---|---|
| Hypermethylated n = 751 | 177 |
| Hypomethylated n = 834 | 162 |
| Unchanged n = 16751 | 17996 |

**Table 3.7: Frequency with which the methylation status of immediately adjacent genes are concordantly changed.** Excluded from this analysis were 18 genes that were both hypermethylated and hypomethylated.

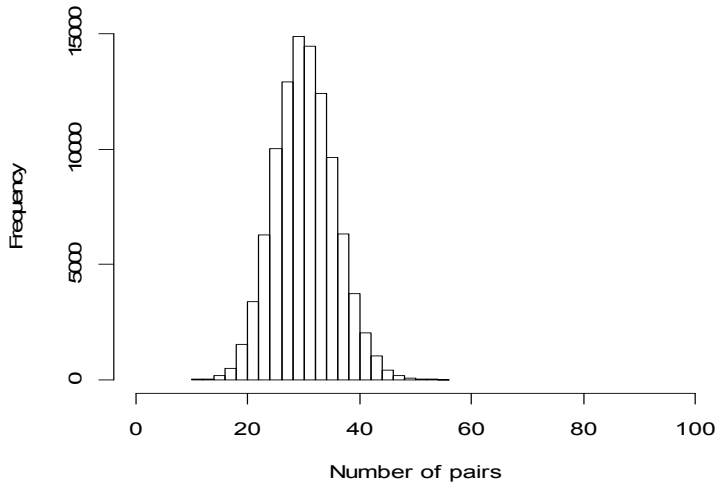



**Figure 3.12:** Histogram shows the distribution of the number of adjacent pairs in 100000 simulated gene sets each containing 751 genes randomly selected from the total of 18354 genes

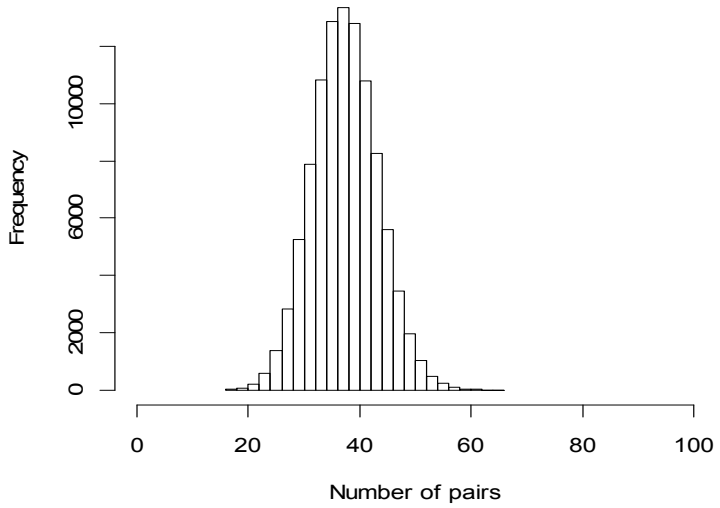



**Figure 3.13:** Histogram shows the distribution of the number of adjacent pairs in 100000 simulated gene sets each containing 834 genes randomly selected from the total of 18354 genes.

| Hypermethylated | |
|---|---|
| Number of pairs | Frequency |
| 11 | 3 |
| 12 | 11 |
| 13 | 7 |
| 14 | 28 |
| 15 | 53 |
| 16 | 111 |
| 17 | 159 |
| 18 | 343 |
| 19 | 602 |
| 20 | 918 |
| 21 | 1385 |
| 22 | 1995 |
| 23 | 2737 |
| 24 | 3528 |
| 25 | 4531 |
| 26 | 5475 |
| 27 | 6083 |
| 28 | 6843 |
| 29 | 7339 |
| 30 | 7531 |
| 31 | 7283 |
| 32 | 7181 |
| 33 | 6575 |
| 34 | 5835 |
| 35 | 5179 |
| 36 | 4445 |
| 37 | 3514 |
| 38 | 2804 |
| 39 | 2170 |
| 40 | 1579 |
| 41 | 1135 |
| 42 | 903 |
| 43 | 597 |
| 44 | 433 |
| 45 | 262 |
| 46 | 150 |
| 47 | 113 |
| 48 | 64 |
| 49 | 44 |
| 50 | 19 |
| 51 | 17 |
| 52 | 5 |
| 53 | 6 |
| 54 | 4 |
| 55 | 1 |
| | |
| | |
| | |
| | |
| | |
| | |

| Hypomethylated | |
|---|---|
| Number of pairs | Frequency |
| 16 | 4 |
| 17 | 4 |
| 18 | 7 |
| 19 | 16 |
| 20 | 33 |
| 21 | 56 |
| 22 | 139 |
| 23 | 216 |
| 24 | 348 |
| 25 | 562 |
| 26 | 810 |
| 27 | 1202 |
| 28 | 1639 |
| 29 | 2274 |
| 30 | 2989 |
| 31 | 3570 |
| 32 | 4317 |
| 33 | 5005 |
| 34 | 5825 |
| 35 | 6327 |
| 36 | 6551 |
| 37 | 6678 |
| 38 | 6671 |
| 39 | 6620 |
| 40 | 6187 |
| 41 | 5692 |
| 42 | 5101 |
| 43 | 4504 |
| 44 | 3747 |
| 45 | 3120 |
| 46 | 2483 |
| 47 | 1905 |
| 48 | 1539 |
| 49 | 1143 |
| 50 | 829 |
| 51 | 600 |
| 52 | 427 |
| 53 | 284 |
| 54 | 189 |
| 55 | 145 |
| 56 | 85 |
| 57 | 64 |
| 58 | 42 |
| 59 | 17 |
| 60 | 15 |
| 61 | 8 |
| 62 | 3 |
| 63 | 1 |
| 64 | 3 |
| 65 | 2 |
| 66 | 2 |

**Table 3.8 Results of simulation analysis performed on 100000 occasions generated using the computer programme R**.

This exercise was repeated for hypomethylated genes. Of the 834 non overlapping genes hypomethylated following EBV infection, 162 were found to be immediately adjacent to another hypomethylated gene (Table 3.7). The distribution of the number of adjacent pairs in 100000 samples each containing 834 genes randomly selected from the total of 18354 genes is shown in Figure 3.13. Given that my simulation exercise suggested that the maximum number of consecutively numbered genes in a 100000 trials was only 66, I conclude that the probability of observing 162 adjacent hypomethylated genes is less than 1 in 100000 (Table.3.8).

Therefore, following infection of GC B cells with EBV, hypermethylated genes are substantially and significantly more likely to be found adjacent to another hypermethylated gene. Likewise, hypomethylated genes were found to be substantially and significantly more likely to be found adjacent to another hypomethylated gene.

### 3.3.2 Methylation changes are enriched at certain chromosomal locations

Having established that hypermethylated and hypomethylated genes are more likely to be found adjacent to each other, I next investigated whether clustering of genes occurred at certain chromosomal locations. To do this, I first counted the number of methylation events which occurred in each cytoband and then calculated the number of methylation events which would have been expected at that cytoband given the overall frequency of methylation events observed on the array. A cytoband was considered to represent a hypermethylation or hypomethylation "hotspot" if there were significantly more methylation events occurring at that location than would have been expected by chance alone; the significance level was set at 1%.

For example, 52 genes on the promoter array map to Chr3q29. 10 genes were observed to be hypermethylated at this chromosomal location. Given that the overall rate of hypermethylation events in this experiment was 0.042 (831/19950), we would have expected 2.2 genes (0.042*52) to be methylated at this location by chance alone. The difference between the observed and expected number provides the basis for significance testing using a chi-square test; in this example, hypermethylation of a gene was 27.33 times more likely to occur at this chromosome location ($p = 0.001$). The hypermethylation hotspots are shown in Table 3.9 and the hypomethylation hotspots are shown in Table 3.10.

| chromosome location | No of genes in cytoband | No of hypermethylated genes in cytoband | Expected | Actual | Actual:Expected | Chi-square (1df) | p value |
|---|---|---|---|---|---|---|---|
| chr9q34.13 | 23 | 7 | 0.98 | 6.019 | > | 36.95 | p<0.001 |
| chr2p23.3 | 66 | 13 | 2.81 | 10.19 | > | 36.88 | p<0.001 |
| chr6q25.3 | 33 | 8 | 1.41 | 6.593 | > | 30.9 | p<0.001 |
| chr3q29 | 52 | 10 | 2.22 | 7.783 | > | 27.33 | p<0.001 |
| chr3q21.1 | 20 | 5 | 0.85 | 4.147 | > | 20.17 | p<0.001 |
| chr11q24.2 | 44 | 8 | 1.88 | 6.124 | > | 19.99 | p<0.001 |
| chr17p13.3 | 70 | 10 | 2.98 | 7.016 | > | 16.49 | p<0.001 |
| chr14q24.3 | 71 | 10 | 3.03 | 6.973 | > | 16.06 | p<0.001 |
| chr1p32.3 | 51 | 8 | 2.17 | 5.826 | > | 15.61 | p<0.001 |
| chr12p11.23 | 10 | 3 | 0.43 | 2.574 | > | 15.54 | p<0.001 |
| chr17p13.2 | 64 | 9 | 2.73 | 6.272 | > | 14.42 | p<0.001 |
| chr7q11.23 | 77 | 10 | 3.28 | 6.717 | > | 13.75 | p<0.001 |
| chr11p15.3 | 11 | 3 | 0.47 | 2.531 | > | 13.66 | p<0.001 |
| chr17q11.2 | 90 | 11 | 3.84 | 7.163 | > | 13.37 | p<0.001 |
| chr9q34.2 | 27 | 5 | 1.15 | 3.849 | > | 12.87 | p<0.001 |
| chr12q13.2 | 61 | 8 | 2.6 | 5.399 | > | 11.21 | p<0.001 |
| chr22q11.21 | 73 | 9 | 3.11 | 5.888 | > | 11.14 | p<0.001 |

**Table 3.9 Clustering of hypermethylated genes by chromosomal location**

| chromosome location | No of genes in cytoband | No of hypomethylated genes in cytoband | Expected | Actual | Actual:Expected | Chi-square (1df) | p value |
|---|---|---|---|---|---|---|---|
| chrXq27.2 | 8 | 5 | 0.37 | 4.626 | > | 57.17 | p<0.001 |
| chr1q44 | 64 | 16 | 2.99 | 13.01 | > | 56.49 | p<0.001 |
| chrXp11.23 | 72 | 17 | 3.37 | 13.63 | > | 55.16 | p<0.001 |
| chr15q11.2 | 40 | 10 | 1.87 | 8.129 | > | 35.31 | p<0.001 |
| chr11q11 | 48 | 11 | 2.25 | 8.754 | > | 34.13 | p<0.001 |
| chr19q13.42 | 91 | 16 | 4.26 | 11.74 | > | 32.39 | p<0.001 |
| chrXp11.22 | 43 | 10 | 2.01 | 7.988 | > | 31.72 | p<0.001 |
| chr11q25 | 13 | 5 | 0.61 | 4.392 | > | 31.71 | p<0.001 |
| chr21q22.2 | 14 | 5 | 0.65 | 4.345 | > | 28.82 | p<0.001 |
| chr2q21.2 | 6 | 3 | 0.28 | 2.719 | > | 26.34 | p<0.001 |
| chr19q13.31 | 51 | 10 | 2.39 | 7.614 | > | 24.3 | p<0.001 |
| chr20p12.1 | 16 | 5 | 0.75 | 4.251 | > | 24.15 | p<0.001 |
| chr10q11.22 | 23 | 6 | 1.08 | 4.924 | > | 22.53 | p<0.001 |
| chr15q13.1 | 7 | 3 | 0.33 | 2.673 | > | 21.81 | p<0.001 |
| chr17q21.2 | 86 | 13 | 4.02 | 8.976 | > | 20.03 | p<0.001 |
| chr1q42.2 | 25 | 6 | 1.17 | 4.83 | > | 19.95 | p<0.001 |
| chr13q21.32 | 1 | 1 | 0.05 | 0.953 | > | 19.42 | p<0.001 |
| chr8p23.2 | 1 | 1 | 0.05 | 0.953 | > | 19.42 | p<0.001 |
| chr18q22.3 | 13 | 4 | 0.61 | 3.392 | > | 18.92 | p<0.001 |
| chr4q35.2 | 13 | 4 | 0.61 | 3.392 | > | 18.92 | p<0.001 |
| chr1p36.21 | 42 | 8 | 1.96 | 6.035 | > | 18.54 | p<0.001 |
| chr2q37.1 | 60 | 10 | 2.81 | 7.193 | > | 18.43 | p<0.001 |
| chr6p25.3 | 8 | 3 | 0.37 | 2.626 | > | 18.42 | p<0.001 |
| chr8p23.1 | 59 | 9 | 2.76 | 6.24 | > | 14.1 | p<0.001 |
| chr12q24.22 | 10 | 3 | 0.47 | 2.532 | > | 13.7 | p<0.001 |
| chr13q21.1 | 5 | 2 | 0.23 | 1.766 | > | 13.33 | p<0.001 |
| chr4p15.31 | 5 | 2 | 0.23 | 1.766 | > | 13.33 | p<0.001 |
| chrXq24 | 42 | 7 | 1.96 | 5.035 | > | 12.9 | p<0.001 |
| chr7q35 | 25 | 5 | 1.17 | 3.83 | > | 12.54 | p<0.001 |
| chr10q26.13 | 34 | 6 | 1.59 | 4.409 | > | 12.22 | p<0.001 |
| chr8p23.3 | 11 | 3 | 0.51 | 2.485 | > | 12 | p<0.001 |

**Table 3.10 Clustering of hypomethylated genes by chromosomal location.**

### 3.3.3 CpG content of cellular genes predict frequency and direction of methylation change

Next, I investigated how the incidence of methylation events varied with the CpG content of the promoter region of a gene. Towards this end I used a classification first described by Weber *et al*., 2007. In this study, genes were categorized according to low CpG (LCP), high CpG (HCP) or intermediate CpG (ICP) content, dependant on their CpG ratio, GC content and length of the CpG-rich region; this classification has since been used by a number of other investigators (Rauch *et al.,* 2009; Ruike *et al.,* 2010).

The analysis was restricted to 14899 genes which were:

- listed on the Affymetrix promoter array (n =19950) and
- listed on the Weber database (n = 15360) and exclusively assigned to either the low, intermediate or high CpG content category on the Weber array (444 genes assigned to two or more categories were excluded for this reason).

Table 3.11 describes the number of genes in each of the CpG content categories, and the frequency with which methylation changes following EBV infection occurred in each category.

| | LCL = GC B | LCL > GC B | LCL < GC B | LCL > and < GC B | **Total** |
|---|---|---|---|---|---|
| LCP | 3737 | 102 | 294 | 0 | **4133** |
| ICP | 1824 | 58 | 118 | 2 | **2002** |
| HCP | 8058 | 465 | 230 | 11 | **8764** |
| **Total** | **13619** | **625** | **642** | **13** | **14899** |

**Table 3.11: Number of genes with increased or decreased methylation that have high, low or intermediate CpG content.** LCL > and < GC B refers to genes that had both hypermethylated and hypomethylated regions.

Table 3.12 describes how the frequency of hypermethylation and hypomethylation events varied according to CpG content using statistical methods described in Section 3.1.8.

| | CpG content | Methylation status changed in EBV infected GC B cells | | Odds ratio | 95% CI | *p* value |
|---|---|---|---|---|---|---|
| | | Yes | No | | | |
| Increased methylation in EBV infected GC B cells | LCP | 102 | 3737 | 1 (referent) | | |
| | ICP | 58 | 1824 | 1.17 | 0.8 to 1.6 | p = 0.3597 |
| | HCP | 465 | 8058 | 2.11 | 1.7 to 2.6 | p = 0.0000 |
| | | | | | | |
| Decreased methylation in EBV infected GC B cells | LCP | 294 | 3737 | 1 (referent) | | |
| | ICP | 118 | 1824 | 0.82 | 0.66 to 1 | p = 0.082 |
| | HCP | 230 | 8058 | 0.36 | 0.3 to 0.4 | p = 0.0000 |

**Table 3.12: Risk of decreased or increased methylation in EBV infected GC B cells in relation to CpG content.** Excluded from this analysis were 13 genes that were both hypermethylated and hypomethylated in different regions.

Compared with those genes with a low CpG content, the risk of methylation following EBV infection was greater for those with an intermediate and high CpG content; the highest risk of EBV associated methylation was seen in those genes with a high CpG content (OR = 2.11, 95% CI 1.7-2.6, p = 0.0000). Compared with those genes with a low CpG content, the risk of a decrease in methylation following EBV infection was significantly less for those genes with an intermediate and high CpG content (OR = 0.36, 95% CI 0.3-0.4, p=0.0000).

### 3.3.4 Baseline gene expression levels predict the frequency and direction of methylation change

I next investigated whether changes in methylation status could be related to the baseline expression of genes in GC B cells.

The analysis was restricted to 17667 genes which were:

- listed on the Affymetrix promoter array (n =19950) and
- listed on the Affymetrix transcriptional array (n = 20765)

Baseline gene expression values were obtained from the intensity scores on the transcriptional arrays

Table 3.13 describes the number of genes in each of the baseline expression categories, and the frequency with which methylation changes occurred in each category.

| Baseline expression GC B cells | Number of genes present on both arrays | Decreased methylation | Increased methylation | No change in methylation |
|---|---|---|---|---|
| **0-24** | 4712 | 346 | 75 | 4291 |
| **25-49** | 2762 | 154 | 67 | 2541 |
| **50-99** | 2433 | 89 | 112 | 2232 |
| **100-199** | 2198 | 23 | 137 | 2038 |
| **200-299** | 1239 | 18 | 85 | 1136 |
| **300-399** | 844 | 13 | 67 | 764 |
| **400-499** | 587 | 8 | 33 | 546 |
| **500-999** | 1453 | 20 | 108 | 1325 |
| **> 1000** | 1439 | 12 | 99 | 1328 |
| **Total** | **17667** | **683** | **783** | **16201** |

**Table 3.13: Baseline expression of genes with a change in methylation status**

Table 3.14 estimates the risk of hypermethylation and hypomethylation events associated with levels of baseline expression using statistical methods described in Section 3.1.8. Compared with genes with a low baseline expression, (0-24), the risk of increased methylation following EBV infection increased with increasing values of baseline expression. This risk reached a plateau when expression values were greater than 200 when the risk was increased approximately 4 fold. Compared with those genes with a low baseline expression, (0-24), the risk of decreased methylation following EBV infection decreased with increasing values of baseline expression. A sustained fall was observed in each level of expression with the greatest decrease found in genes with a baseline expression value greater than 1000 (OR 0.11 95% CI, 0.1-0.2, p = 0.0000).

| | Baseline gene intensity scores in GC B cells | Methylation status changed in EBV infected GC B cells | | Odds ratio | 95% CI | *p* value |
|---|---|---|---|---|---|---|
| | | Yes | No | | | |
| Risk of increased methylation in EBV infected GC B cells | 0-24 | 75 | 4291 | **1 (referent)** | | |
| | 25-49 | 67 | 2541 | **1.51** | 1.1 to 2.1 | p = 0.0148 |
| | 50-99 | 112 | 2232 | **2.87** | 2.1 to 3.9 | p = 0.0000 |
| | 100-199 | 137 | 2038 | **3.84** | 2.9 to 5.1 | p = 0.0000 |
| | 200-299 | 85 | 1136 | **4.28** | 3.1 to 5.9 | p = 0.0000 |
| | 300-399 | 67 | 764 | **5.01** | 3.6 to 7.0 | p = 0.0000 |
| | 400-499 | 33 | 546 | **3.46** | 2.3 to 5.3 | p = 0.0000 |
| | 500-999 | 108 | 1325 | **4.67** | 3.5 to 6.3 | p = 0.0000 |
| | > 1000 | 99 | 1328 | **4.27** | 3.1 to 5.8 | p = 0.0000 |
| | | | | | | |
| Risk of decreased methylation in EBV infected GC B cells | 0-24 | 346 | 4291 | **1 (referent)** | | |
| | 25-49 | 154 | 2541 | **0.75** | 0.6 to 0.9 | p = 0.0042 |
| | 50-99 | 89 | 2232 | **0.49** | 0.4 to 0.6 | p = 0.0000 |
| | 100-199 | 23 | 2038 | **0.14** | 0.1 to 0.2 | p = 0.0000 |
| | 200-299 | 18 | 1136 | **0.20** | 0.1 to 0.3 | p = 0.0000 |
| | 300-399 | 13 | 764 | **0.21** | 0.1 to 0.4 | p = 0.0000 |
| | 400-499 | 8 | 546 | **0.18** | 0.1 to 0.4 | p = 0.0000 |
| | 500-999 | 20 | 1325 | **0.19** | 0.1 to 0.3 | p = 0.0000 |
| | > 1000 | 12 | 1328 | **0.11** | 0.1 to 0.2 | p = 0.0000 |

**Table 3.14: Baseline gene intensity scores in uninfected GC B cells predict methylation changes in GC B cells.**

### 3.3.5 CpG content and baseline expression scores are independent predictors of methylation change

Given that both CpG content and baseline expression appear to predict changes in methylation status, I next investigated, whether these were independent effects.

The analysis was restricted to 13997 genes which were:

- listed on the Affymetrix promoter array (n =19950)
- listed on the Affymetrix transcriptional array (n = 20765)
- listed on the Weber database (n = 15360) and exclusively assigned to either the low, intermediate or high CpG content category on the Weber array (444 genes assigned to two or more categories were excluded for this reason).

I first stratified the dataset by CpG content and then by baseline gene expression. Table 3.15 shows that within each CpG content category, the risk of increased methylation following EBV infection increased with increasing baseline expression. Table 3.16 shows that within each CpG content category the risk of decreased methylation following EBV infection is decreased with increasing baseline expression. Therefore, this stratified analysis suggests that both CpG content and baseline expression independently predict changes in methylation status following EBV infection.

| CpG island content | Baseline gene intensity scores in GC B cells | Methylation status changed in EBV infected GC B cells | | Odds ratio | 95% CI | *p* value |
|---|---|---|---|---|---|---|
| | | Increased | No change | | | |
| LCP | 0-24 | 23 | 1459 | **1 (referent)** | | |
| | 25-49 | 10 | 690 | **0.91** | 0.4 to 1.9 | p = 0.8255 |
| | 50-99 | 20 | 405 | **3.13** | 1.7 to 5.7 | p = 0.0001 |
| | 100-199 | 10 | 239 | **2.65** | 1.2 to 5.6 | p = 0.0182 |
| | 200-499 | 17 | 210 | **5.14** | 2.7 to 9.8 | p = 0.0000 |
| | > 500 | 19 | 232 | **5.2** | 2.8 to 9.7 | p = 0.0000 |
| | | | | | | |
| ICP | 0-24 | 6 | 592 | **0.64** | 0.3 to 1.6 | p = 0.3341 |
| | 25-49 | 6 | 313 | **1.21** | 0.5 to 3.0 | p = 0.6720 |
| | 50-99 | 12 | 261 | **2.92** | 1.4 to 5.9 | p = 0.0020 |
| | 100-199 | 10 | 191 | **3.32** | 1.6 to 7.1 | p = 0.0010 |
| | 200-499 | 10 | 179 | **3.54** | 1.7 to 7.6 | p = 0.0005 |
| | > 500 | 11 | 181 | **3.86** | 1.8 to 8.0 | p = 0.0001 |
| | | | | | | |
| HCP | 0-24 | 29 | 1303 | **1.41** | 0.8 to 2.5 | p = 0.2188 |
| | 25-49 | 33 | 961 | **2.18** | 1.3 to 3.7 | p = 0.0037 |
| | 50-99 | 60 | 1075 | **3.54** | 2.2 to 5.8 | p = 0.0000 |
| | 100-199 | 76 | 1160 | **4.16** | 2.6 to 6.7 | p = 0.0000 |
| | 200-499 | 116 | 1574 | **4.68** | 3.0 to 7.4 | p = 0.0000 |
| | > 500 | 146 | 1812 | **5.11** | 3.3 to 8.0 | p = 0.0000 |

**Table 3.15 Baseline gene intensity scores in uninfected GC B cells independently predict hypermethylation events in EBV infected GC B cells.**

| CpG island content | Baseline gene intensity scores in GC B cells | Methylation status change in EBV infected GC B cells | | Odds ratio | 95% CI | *p* value |
|---|---|---|---|---|---|---|
| | | Decreased | No change | | | |
| LCP | 0-24 | 143 | 1459 | **1 (referent)** | | |
| | 25-49 | 57 | 690 | **0.84** | 0.6 to 1.2 | p = 0.2946 |
| | 50-99 | 22 | 405 | **0.55** | 0.3 to 0.9 | p = 0.0112 |
| | 100-199 | 5 | 239 | **0.21** | 0.1 to 0.5 | p = 0.0002 |
| | 200-499 | 3 | 210 | **0.15** | 0.05 to 0.5 | p = 0.0001 |
| | > 500 | 3 | 232 | **0.13** | 0.04 to 0.4 | p = 0.0000 |
| | | | | | | |
| ICP | 0-24 | 48 | 592 | **0.83** | 0.6 to 1.2 | p = 0.2745 |
| | 25-49 | 24 | 313 | **0.78** | 0.5 to 1.2 | p = 0.2831 |
| | 50-99 | 12 | 261 | **0.47** | 0.3 to 0.9 | p = 0.0119 |
| | 100-199 | 3 | 191 | **0.16** | 0.05 to 0.5 | p = 0.0003 |
| | 200-499 | 2 | 179 | **0.11** | 0.03 to 0.5 | p = 0.0002 |
| | > 500 | 4 | 181 | **0.23** | 0.1 to 0.6 | p = 0.0015 |
| | | | | | | |
| HCP | 0-24 | 83 | 1303 | **0.65** | 0.5 to 0.9 | p = 0.0024 |
| | 25-49 | 36 | 961 | **0.38** | 0.3 to 0.6 | p = 0.0000 |
| | 50-99 | 43 | 1075 | **0.41** | 0.3 to 0.6 | p = 0.0000 |
| | 100-199 | 15 | 1160 | **0.13** | 0.1 to 0.2 | p = 0.0000 |
| | 200-499 | 23 | 1574 | **0.15** | 0.1 to 0.2 | p = 0.0000 |
| | > 500 | 20 | 1812 | **0.11** | 0.1 to 0.2 | p = 0.0000 |

**Table 3.16: Baseline gene intensity scores in uninfected GC B cells independently predict demethylation events in EBV infected GC B cells**.

### 3.3.6 EBV induced methylation changes are unreliable predictors of EBV associated transcriptional change

I next investigated how often methylation changes observed in my EBV-GC B cell arrays were followed by transcriptional changes. Towards this end, I used the EBV-GC B cell methylation array and the transcriptional array (re-analysed on this occasion using a 1% FDR cut-off and 1.5 fold change cut off were used).

The analysis was restricted to 17667 genes which were:

- listed on the Affymetrix promoter array (n =19950)
- listed on the Affymetrix transcriptional array (n = 20765)

I examined the relationship between the direction of the methylation change and the direction of the transcriptional change. Table 3.17 suggests that, compared with those genes in which there was no change in methylation status following EBV infection, transcriptional up-regulation was significantly more likely to occur in those genes with increased methylation (OR 1.54, 95% CI 1.3 to 1.8, *p*=1.0E-6). However, transcriptional down-regulation was not significantly associated with an increase in methylation (OR 1.3, 95% CI 0.96 to 1.4, *p*=0.135) (Table 3.18).

| | | Number of genes | Number of genes with increased expression following EBV infection of GC B cells | Odds ratio | 95% CI | *p* value |
|---|---|---|---|---|---|---|
| Change in methylation status | No change | 16184 | 2542 | 1 (referent) | | |
| | Increased | 783 | 175 | 1.54 | 1.3 to 1.8 | *p*=1.0E-6 |

Table 3.17: An increase in methylation is significantly associated with an increase in expression.

| | | Number of genes | Number of genes with decreased expression following EBV infection of GC B cells | Odds ratio | 95% CI | *p* value |
|---|---|---|---|---|---|---|
| Change in methylation status | No change | 16184 | 2393 | 1 (referent) | | |
| | increased | 783 | 131 | 1.3 | 0.96 to 1.4 | *p*=0.135 |

Table 3.18: An increase in methylation is not significantly associated with a decrease in expression.

Tables 3.19 shows that, compared with those genes in which there was no change in methylation status following EBV infection, transcriptional up-regulation and was less likely to occur in those genes that were hypomethylated following infection (OR 0.3, 95% CI 0.22 to 0.43, *P* =0.001). This was also the case for those transcriptionally down-regulated genes (OR 0.57, 95% CI 0.44 to 0.74, p=1.9E-5).

| | | Number of genes | Number of genes with increased expression following EBV infection of GC B cells | Odds ratio | 95% CI | *p* value |
|---|---|---|---|---|---|---|
| Change in methylation status | No change | 16184 | 2542 | 1 (referent) | | |
| | Decreased | 683 | 37 | 0.3 | 0.22 to 0.43 | *P* =0.001 |

Table 3.19: A decrease in methylation is not associated with an increase in expression.

| | | Number of genes | Number of genes with decreased expression following EBV infection of GC B cells | Odds ratio | 95% CI | *p* value |
|---|---|---|---|---|---|---|
| Change in methylation status | No change | 16184 | 2393 | 1 (referent) | | |
| | Decreased | 683 | 66 | 0.57 | 0.44 to 0.74 | *p*=1.9E-5 |

Table 3.20: A decrease in methylation status is not associated with a decrease in expression.

**Summary**

In this section, I have shown that following EBV infection, hypermethylated and hypomethylated genes are frequently found immediately adjacent to concordantly methylated genes. I have also shown that methylation hotspots can be found at certain chromosomal locations. Baseline gene expression and cellular gene CpG content were found to be independent predictors of the frequency and direction of methylation changes. These global analyses also suggested that methylation changes are likely to be followed by an increased incidence of transcriptional changes. However, there was no clear correspondence between the direction of the methylation change and the direction of the transcriptional change. Increased methylation was significantly associated with transcriptional up-regulation but not down-regulation, However, EBV associated hypomethylation was significantly less likely to be associated with transcriptional change.

## 3.4 Ontological characterisation of EBV associated methylation changes

In this section, I first report the ontological association of those genes found to be either hypermethylated or hypomethylated in EBV infected GC B cells. I next present the results of a literature search designed to identify those genes reported to be regulated by methylation in HL, and report how often these were also found to be methylated in EBV infected GC B cells.

### 3.4.1 Genes hypomethylated by EBV in GC B cells are enriched for those involved in G-protein coupled signalling and for cancer testis antigens

The ontology of those genes found to have had a change in methylation status following EBV infection was characterised using the PANTHER classification system. This system uses the gene ontology TM for classifications by molecular function, biological process and cellular component. The referent gene list used in these analyses comprised all genes listed on the promoter array. Ontological profiling of genes which were hypermethylated following EBV infection revealed no significant associations. However, genes with reduced methylation following EBV infection of GC B cells were significantly enriched for a number of biological processes listed in Table 3.21.

| Biological Process | Number of genes with no change in methylation status | Number of genes with decreased methylation | expected | *p* value (Bonferroni correction applied) |
|---|---|---|---|---|
| Chemosensory perception | 149 | 38 | 7.19 | 4.76E-14 |
| Olfaction | 142 | 37 | 6.85 | 8.31E-14 |
| Signal transduction | 2739 | 204 | 132.15 | 1.34E-09 |
| G-protein mediated signalling | 653 | 73 | 31.51 | 1.09E-08 |
| Sensory perception | 389 | 49 | 18.77 | 7.49E-08 |
| Cell surface receptor mediated signal transduction | 1312 | 113 | 63.3 | 2.47E-07 |
| Cell adhesion-mediated signalling | 310 | 39 | 14.96 | 2.16E-05 |
| Cell communication | 988 | 79 | 47.67 | 0.00141 |
| Oncogenesis | 330 | 33 | 15.92 | 0.00302 |
| Ectoderm development | 545 | 48 | 26.3 | 0.00949 |
| Receptor protein serine/threonine kinase signalling pathway | 26 | 8 | 1.25 | 0.00969 |

**Table 3.21: Biological processes significantly enriched among genes with reduced methylation following EBV infection of GC B cells.**

Of course, as with all ontological classifications, one gene can be assigned to more than one biological process. For example, consider three of the biological processes listed in Table 3.21; sensory perception, chemosensory perception and olfaction. These three processes contributed to the ontological profiling of hypomethylation, 49, 38 and 37 EBV demethylated genes, respectively (highlighted in blue in Table 3.19). Figure 3.14 reveals that most of these hypomethylated genes are common to all three processes.

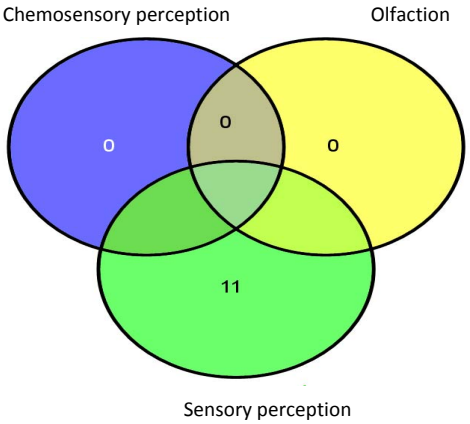



**Figure 3.14: Overlap of hypomethylated genes involved in sensory perception, chemosensory perception and olfaction.**

Of the 49 unique genes contributing to these analyses of perception, 39 were also found to be involved in G-protein mediated signalling, another biological process significantly over-represented in EBV demethylated genes (Figure 3.15).

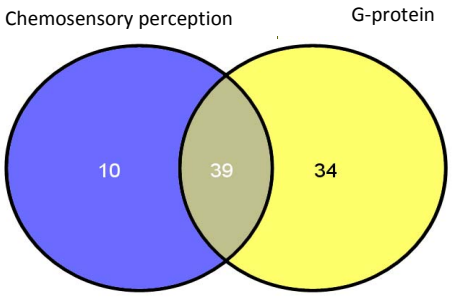



**Figure 3.15: Overlap of genes involved in perception and G protein mediated signalling.**

If we now focus on genes involved in G-protein mediated signalling, we see that all of these genes (n = 73) were also involved in cell surface receptor mediated signal transduction, another biological process significantly over-represented in EBV demethylated genes (Figure 3.16)

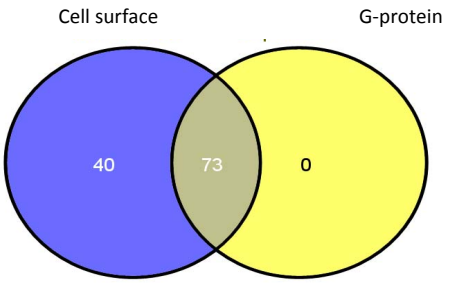



**Figure 3.16: Genes common to cell surface receptor mediated signalling and G protein mediated signalling.**

All of the 73 genes involved in G-protein mediated signalling are also included in the 204 signal transduction genes referred to in Table 3.21.

Therefore, it would seem that EBV-associated demethylation of G-protein mediated signalling genes alone can explain many of the ontological associations observed in this dataset. Given that we have already shown that genes with a low CpG content are significantly more likely to be hypomethylated following EBV infection, it is noteworthy that genes involved in G protein mediated signalling are significantly enriched for those with a low CpG content (p = 3.98E-132).

PANTHER gene database does not of course identify all gene subsets of potential interest. For example it does not include a listing of cancer testis antigens (CTA), which are often hypomethylated in cancer: However, I was able to identify using the CTDatabase compiled by the Ludwig Institute and supplemented by a search of the NCBI database, 232 CTA. Of these, 30 CTA were found to be hypomethylated following EBV infection of GC B cells (OR = 5.5, 95% CI; 3.7 to 8.1; p = 0.0000). Some of the CTA, for example, the SSX genes and others from the same genomic location (CT45A1) have been shown to be over-expressed in HL (Chen *et al.,* 2010).

| Changes in the methylation status of Cancer-Testis Antigens (CTA) following infection of GC B cells with EBV | |
|---|---|
| | Number of CTA |
| Hypomethylated | 30 |
| Unchanged | 180 |
| Hypermethylated | 0 |
| Not on promoter array | 22 |

**Table 3.22: Cancer testis-antigens hypomethylated following EBV infection of GC B cells.**

### 3.4.2 Tumour suppressor genes known to be silenced by methylation in HL are not hypermethylated by EBV in GC B cells

The germinal centre B cell is considered the progenitor cell for HL. Ideally I would compare the results of my methylation array on EBV infected GC B cells with a comparable genome wide analysis of the methylation changes in HL. However no such analysis has been undertaken. In order to generate a comprehensive list of those genes known to be methylated in HL, I performed a literature search using the terms "Hodgkin's lymphoma" and "DNA methylation". This search identified only 21 genes which were known to be hypermethylated in HL (Table 3.23); one of these TP73 was also found to be methylated by EBV on my array. To date, no genes have been reported to be hypomethylated in HL.

| Gene symbol | Pubmed identifier | | |
|---|---|---|---|
| DAPK2 | 19609235 | | |
| POU2AF1 | 18256685 | 16304050 | |
| NFATC1 | 18156209 | | |
| TNFRSF8 | 17965727 | | |
| DLC1 | 17965626 | | |
| PCDH10 | 17341268 | | |
| SYK | 16304050 | | |
| CD79B | 16304050 | | |
| IGH@ | 15284123 | | |
| CHEK2 | 15153943 | | |
| RASSF1 | 14961078 | | |
| CDKN2C | 14645011 | | |
| DAPK1 | 14595709 | 10361133 | |
| CDKN2A | 12213729 | 10616196 | 9473234 |
| CDKN2B | 12213729 | 9473234 | |
| BCL2L11 | 19557159 | | |
| MSC | 18368067 | | |
| BMP6 | 17575215 | | |
| CDH1 | 17214905 | | |
| FHIT | 15384174 | | |
| TP73 | 10416592 | | |
| OPCML | 18714356 | | |

**Table 3.23: Genes hypermethylated in HL from the literature.** Pubmed identifiers are also recorded

**Summary**

In this section, I have shown that genes hypomethylated following infection of GC B cells with EBV are enriched for those involved in G protein mediated signalling. These genes are also enriched for those with a low CpG content, and therefore as I have shown in the previous section, are more susceptible to hypomethylation. Finally, I could find no compelling evidence to suggest that EBV-associated methylation changes in GC B cells are recapitulated in HL. However, this might be better interpreted as the absence of evidence rather than evidence of absence.

Chapter 4

# The impact of EBV and its latent genes on the expression of proteins regulating DNA and arginine methylation

## Aims of Chapter 4.

1. To profile the expression of the DNA methyltransferases, DNMT1, DNMT3A and DNMT3B
    a) following the infection of GC B cells with EBV
    b) in different B cell subsets
    c) in a panel of HL cell lines
2. To determine if expression of one or more of the DNA methyltransferases is regulated by the EBV latent genes, LMP1, LMP2A and EBNA1
3. To profile the expression of the protein arginine methyltransferases, CARM1, PRMT1, PRMT5, and the peptidylarginine deiminase, PADI4,
    a) following the infection of GC B cells with EBV
    b) in a panel of HL cell lines
4. To determine if the expression of those enzymes modulating arginine methylation is regulated by the EBV latent gene LMP1.

In the first section of this chapter, I show how the expression of DNA methyltransferases, DNMT1, DNMT3A and DNMT3B changes at different time-points following EBV infection of GC B cells. I next profile the expression of these proteins across different B cell subsets and in a panel of HL cell lines. Finally, I investigate whether the expression of one or more of the DNMTs is modulated by the EBV latent genes, LMP1, LMP2A, or EBNA1.

## 4.1 EBV infection of GC B cells modulates the expression of DNMTs

The microarray performed on GC B cells and LCLs six week post infection revealed that the expression of DNMT3A was up-regulated more than 8 fold in the LCLs compared to GC B cells and that of DNMT1 and DNMT3B reduced (Table 4.1).

| NCBI gene symbol | Probe set ID | Fold change |
|---|---|---|
| DNMT1 | 201697_s_at | -1.8 |
| DNMT3B | 220668_s_at | -2.9 |
| DNMT3A | 222640_at | 8.7 |

**Table 4.1: Microarray analysis results for the DNMTs.** SAM was performed with an FDR of 5%.

I next validated these changes at the RNA level using Q RT-PCR, and at the protein level using western blotting. Q RT-PCR analysis of DNMT1, DNMT3A and DNMT3B was carried out on RNA harvested six weeks post infection on three newly established LCLs derived from the GC B cells of three different donors. RNA extraction, cDNA synthesis and Q RT-PCR analysis was carried out using the method described in Section 2.2.1, 2.3 and 2.4. Taqman assays which cover the known isoforms of the DNMTs are listed in Section 2.4.4 (page 57).

Compared with uninfected GC B cells, the expression of DNMT1 and DNMT3B was significantly down-regulated in all three LCLs (Figure 4.1A and B). On the other hand, there was a substantial increase (9-15 fold) in DNMT3A expression (Figure 4.1C). Western blotting confirmed that both isoforms 1 and 2 of DNMT1 were down-regulated in LCLs compared to GC B cells. DNMT3B was down-regulated in all LCLs while DNMT3A was up-regulated at the protein level.

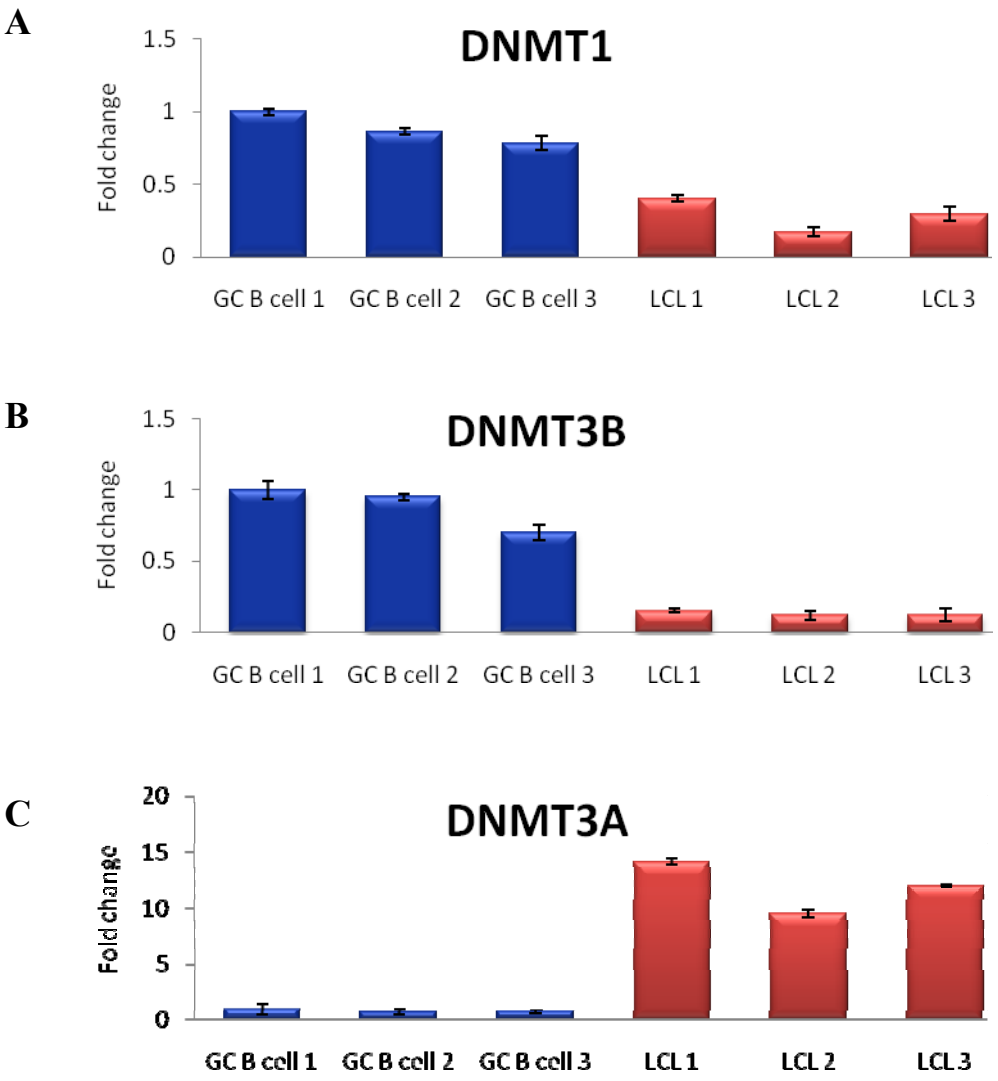



**Figure 4.1 A-C: DNMT1, DNMT3A and DNMT3B mRNA expression in LCLs.** Blue bars represent three different patient GC B cell samples and red bars represent each of the LCLs. The GC B cells with the highest expression of the gene in question served as the reference with which other GC B cell samples and LCLs were compared. Assays were carried out in triplicate and results are presented as $2^{-\Delta\Delta CT}$ values. Standard error bars show the variation in relative DNMT expression across three independent experiments.

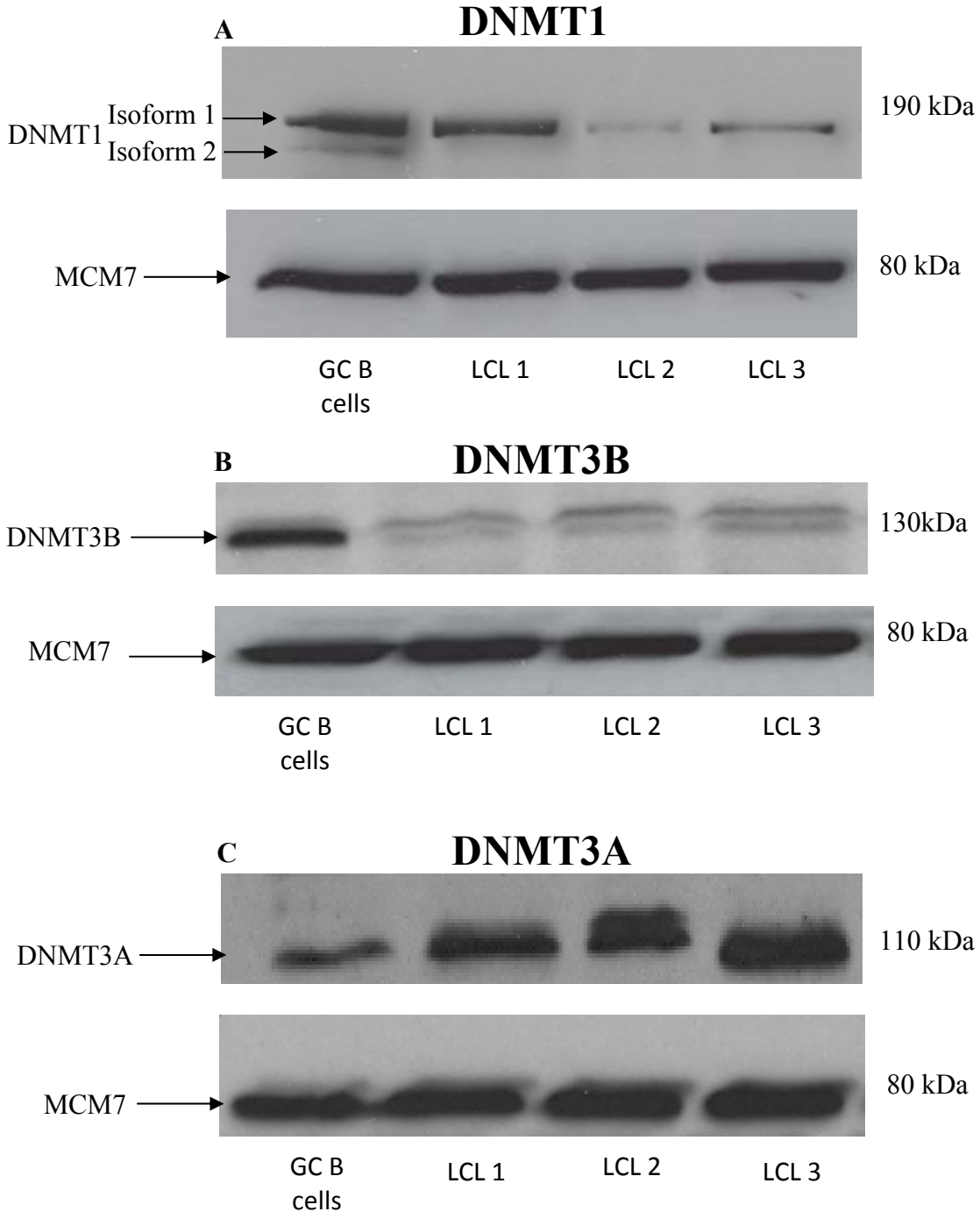



**Figure 4.2 A-C: Protein expression of DNMTs in LCLs compared to GC B cells:** **A.** DNMT1 expression detected using 15 µg of protein. **B.** DNMT3B expression detected using 30 µg of protein. **C.** DNMT3A expression detected using 45µg of protein. MCM7 was used as a loading control for each western blot. The molecular weight for each protein is shown on the right.

## 4.2 Changes in the expression of the DNMTs occur soon after onset of EBV infection

RNA and DNA was also taken from the freshly infected GC B cells 24 hrs, 48 hrs, 72 hrs, 96 hrs, 1 wk, 2 wks and 3 wks post infection. These samples were used to assay the kinetics of EBV-induced de-regulation of the DNMTs. Figure 4.3A shows that DNMT1 was down-regulated within 48 hours of infection, and continued to remain low throughout the time course. DNMT3B was down-regulated within 24 hours (Figure 4.3B). DNMT3A expression was first raised 72 hours post infection, and increased gradually thereafter (Figure 4.3C).

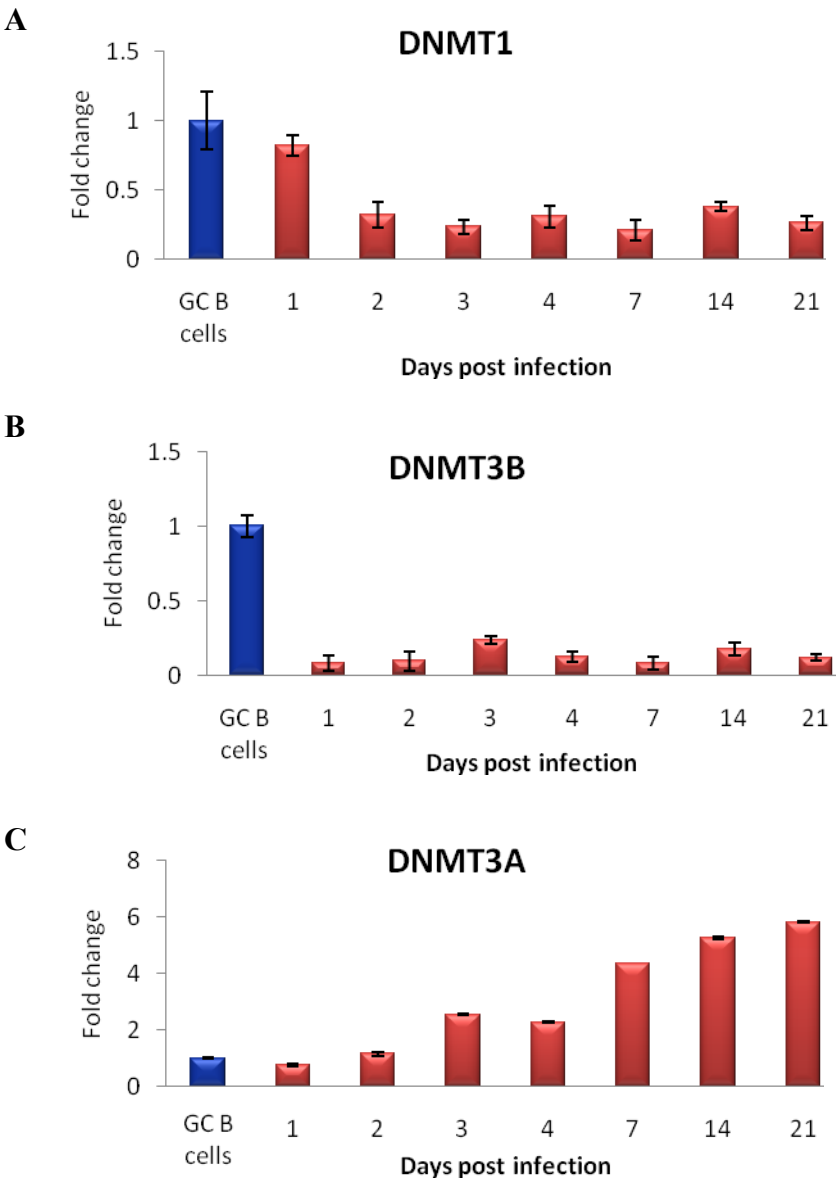



**Figure 4.3 A-C: Kinetics of DNMT1, DNMT3A and DNMT3B expression in LCLs by Q-RT PCR.** The expression at each time point was compared to that of the GC B cells. Assays were carried out in triplicate and results are presented as $2^{-\Delta\Delta CT}$ values. Results are representative of an experiment performed on three different LCLs. Standard error bars show the variation of DNMT expression across the three assays.

## 4.3 DNMT expression varies across B cell subsets

For these experiments, I used naive B cell, centroblast, centrocyte and memory B cell subsets which had been isolated from tonsillar tissue by members of our laboratory. The isolation procedure is now described in brief. $CD77^+$ centroblasts were enriched by magnetic separation with α-CD77 antibody and α-IgM microbeads using LS columns. The flow-through of the magnetic separation containing CD77 depleted mononuclear cells was incubated with α-CD10-Phycoerythrin (PE) and α-PE microbeads followed by magnetic enrichment on LS columns. The eluate contained $CD10^+/CD77^-$ centrocytes. Naïve B cells were isolated indirectly by depletion of all non-naïve cells using a cocktail of antibodies to CD10, CD2, CD16, CD27, CD36, CD43 and CD235a. Memory B cells were first purified by negative depletion using antibodies against CD2, CD14, CD16, CD36, CD43 and CD235a, and then by positive selection with CD27 microbeads. The relative expression of the DNMTs in each B cell subset was analysed.

Q RT-PCR showed that DNMT1 expression was highest in centroblasts and lowest in naive B cells, with centrocytes and memory cells having similar levels of expression (Figure 4.4A). DNMT3A expression was highest in the naive cells and lowest in the centroblasts, with centrocytes and memory cells again having similar levels of expression (Figure 4.4B). DNMT3B expression was highest in centroblasts and lowest in memory B cells (Figure 4.4C).

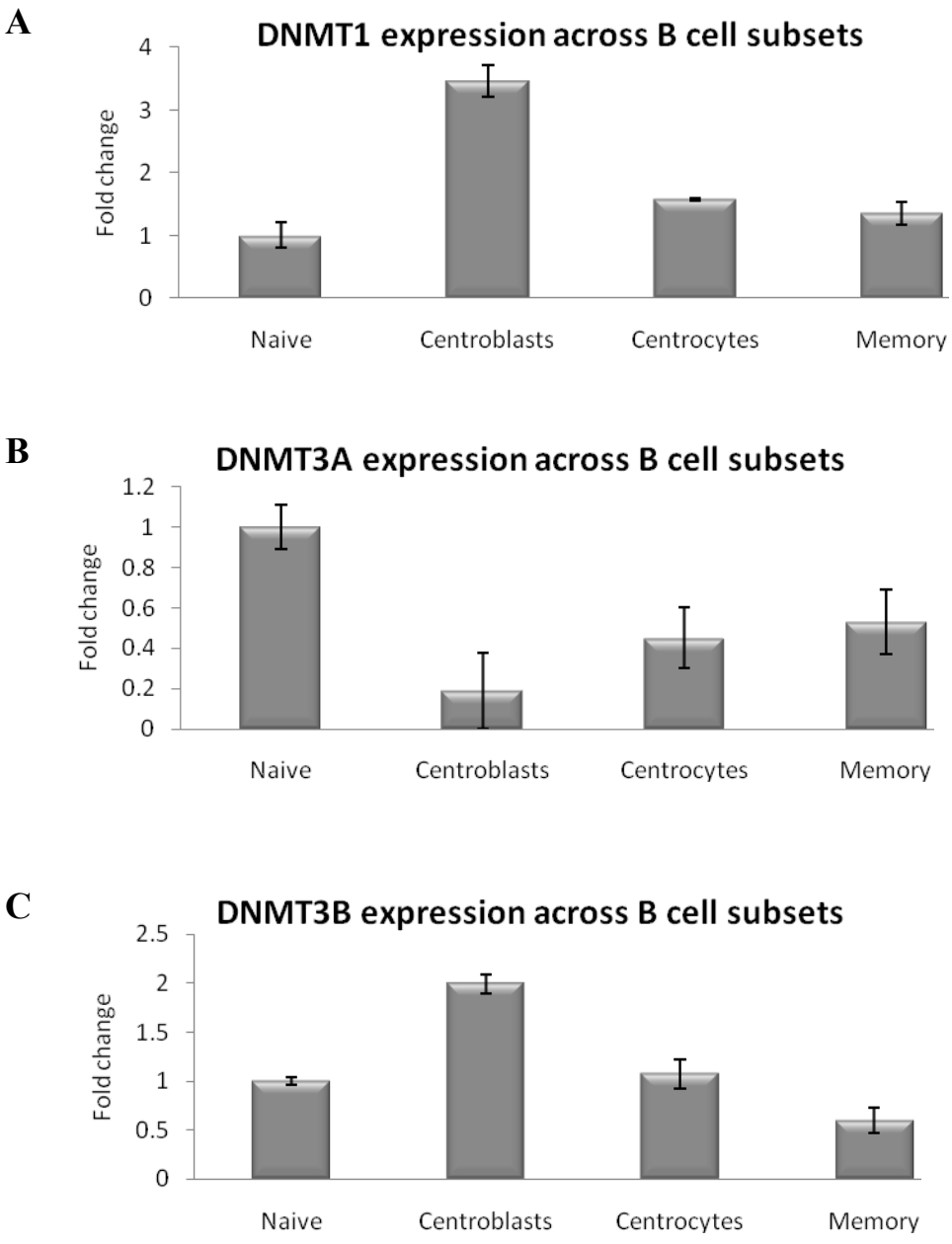
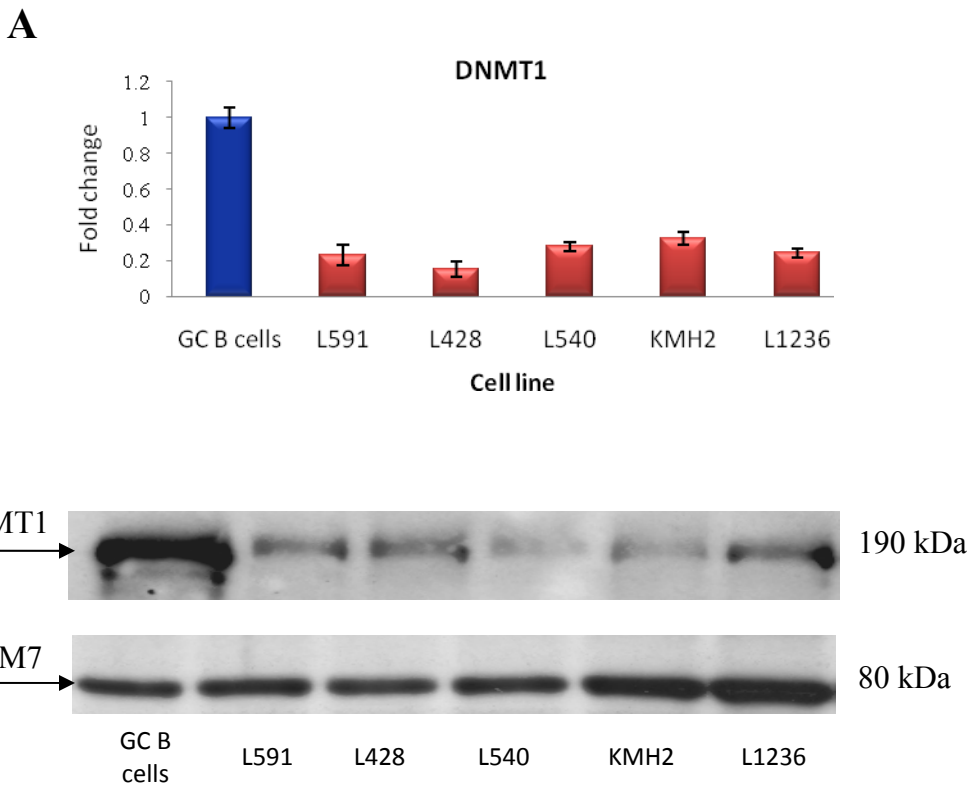


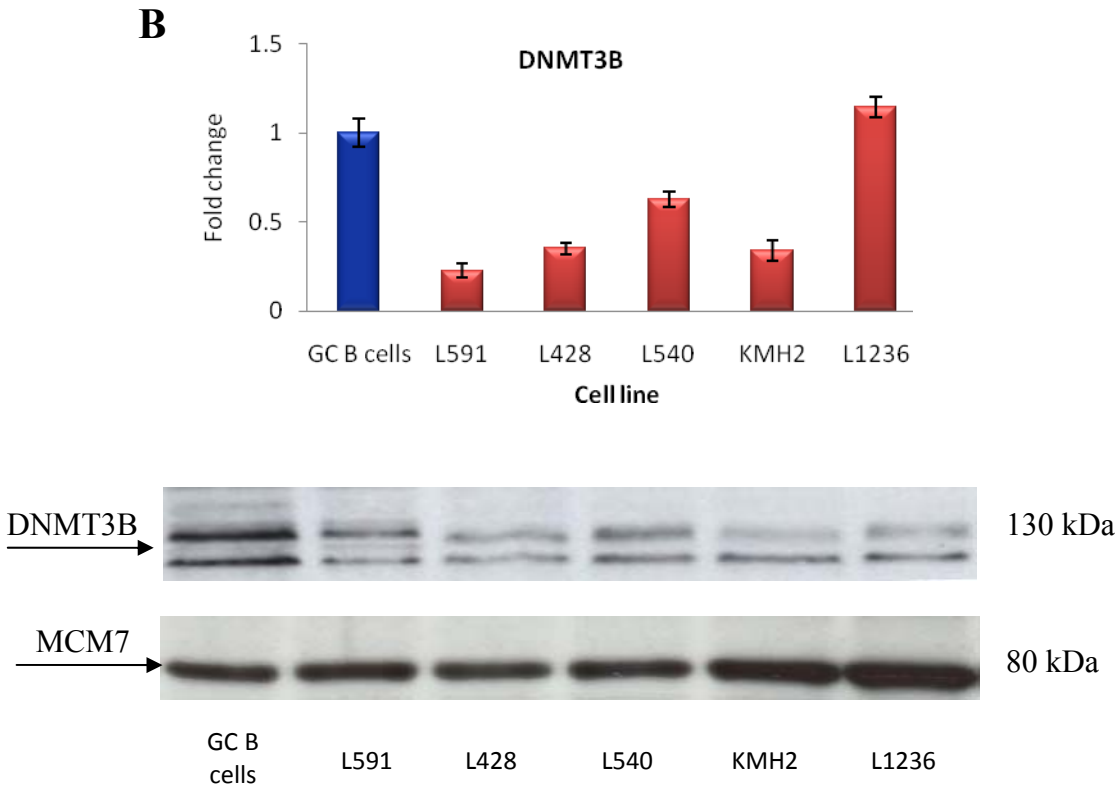

**Figure 4.4 A-C: DNMT expression across B cell subsets.** DNMT1, DNMT3A and DNMT3B expression was assayed across different B cell subsets using Taqman Q RT-PCR. Assays were carried out in triplicate and results are presented as $2^{-\Delta\Delta CT}$ values using naive B cells as the reference sample. Standard error bars show the variation of DNMT expression across the three assays.

## 4.4 DNMTs are differentially expressed in HL cell lines

I next explored how the expression of the DNMTs at the transcriptional and protein level in a panel of five HL cell lines differed from that of GC B cells isolated from different donors The origin and culturing of these cell lines is described in section 2.1.

Compared with GC B cells, Q RT-PCR revealed the down-regulation of DNMT1in all HL cell lines; the down-regulation of DNMT3B in all cell lines with the exception of L1236; and the up-regulation of DNMT3A in all cell lines with the exception of L428. However, western blotting confirmed the down-regulation of DNMT1 and DNTM3B and the up-regulation of DNMT3A in all HL cell lines at the protein level (figure 4.5A-C).


A
DNMT1
Fold change
1.2
1
0.8
0.6
0.4
0.2
0
GC B cells
L591
L428
L540
KMH2
L1236
Cell line
DNMT1
190 kDa
MCM7
80 kDa
GC B cells
L591
L428
L540
KMH2
L1236


B
DNMT3B
Fold change
1.5
1
0.5
0
GC B cells
L591
L428
L540
KMH2
L1236
Cell line
DNMT3B
130 kDa
MCM7
80 kDa
GC B cells
L591
L428
L540
KMH2
L1236

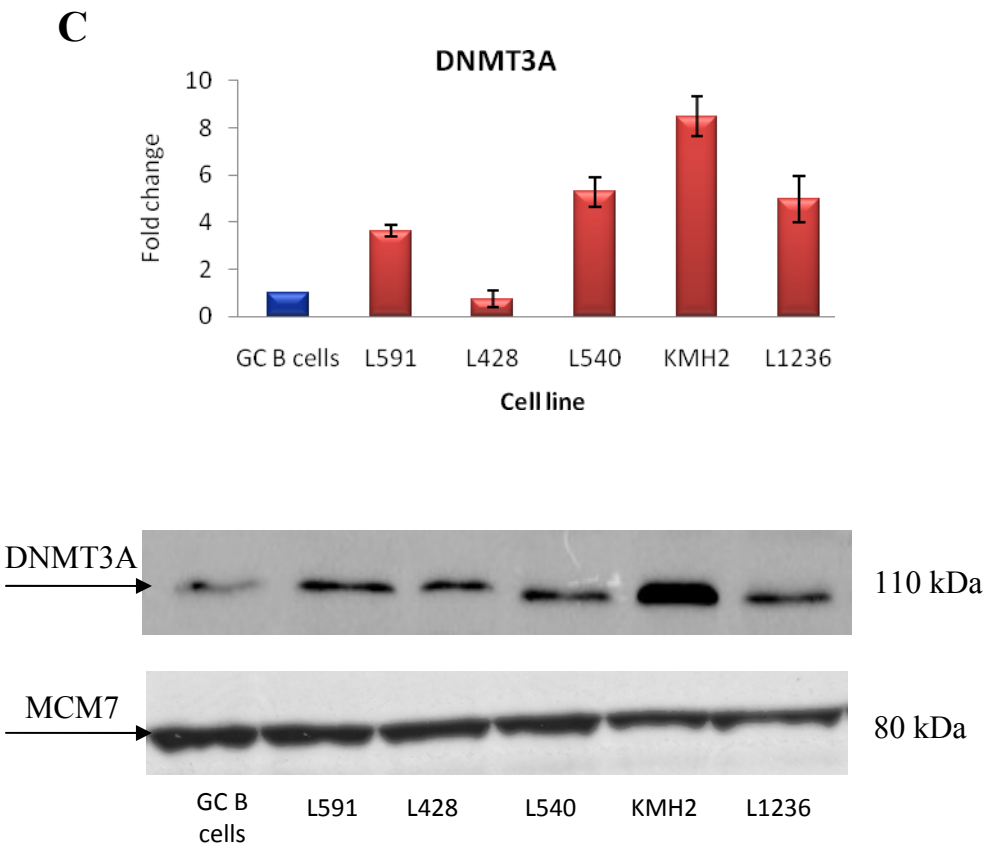



**Figure 4.5 A-C: Q RT-PCR and western blotting analysis of DNMT expression in a panel of five HL cell lines.** For Q RT -PCR data, the blue bar represents GC B cells and each of the red bars represents a different HL cell line. For each DNMT, the expression in HL cell lines was compared to that of GC B cells. All Q RT-PCR samples were analysed in triplicate and are presented as $2^{-\Delta\Delta CT}$ values compared with GC B cells. Protein levels were measured using western blotting with MCM7 used as a loading control. **A.** DNMT1 expression in a panel of cell lines analysed by Q RT -PCR and WB (15 µg protein). **B**. DNMT3B expression in a panel of cell lines analysed by Q RT-PCR and WB (30 µg protein). **C**. DNMT3A expression in a panel of cell lines analysed by Q RT-PCR and WB (45 µg protein). The molecular weight of each of the proteins is shown.

## 4.5 DNMT1 is down-regulated by LMP1 but not by EBNA1 or LMP2A in GC B cells

Having shown that EBV can de-regulate the DNMTs, I next investigated whether their expression was regulated by one or more of the EBV latent genes, LMP1, LMP2A or EBNA1.

I first investigated whether LMP1 down-regulated DNMT1 in GC B cells using RNA from an experiment in which GC B cells obtained from different patients had been transfected either with a pSG5-LMP1expression vector or with a pSG5 vector control (Vockerodt *et al*., 2008). I was able to confirm using Q-RT PCR and RNA isolated 24 hours post transfection that LMP1 down-regulated DNMT1 in these GC B cells (Figure 4.6). I next investigated whether DNMT1 was regulated by either LMP2A or EBNA1. For these experiments I used RNA which had been obtained from GC B cells transfected with LMP2A and EBNA1. These transfections had been performed by Dr. Vockerodt and Dr. Kapatai, respectively, according to methods previously described (Vockerodt *et al*., 2008). Neither LMP2A nor EBNA1 modulated the expression of DNMT1 in GC B cells (Figure 4.7).

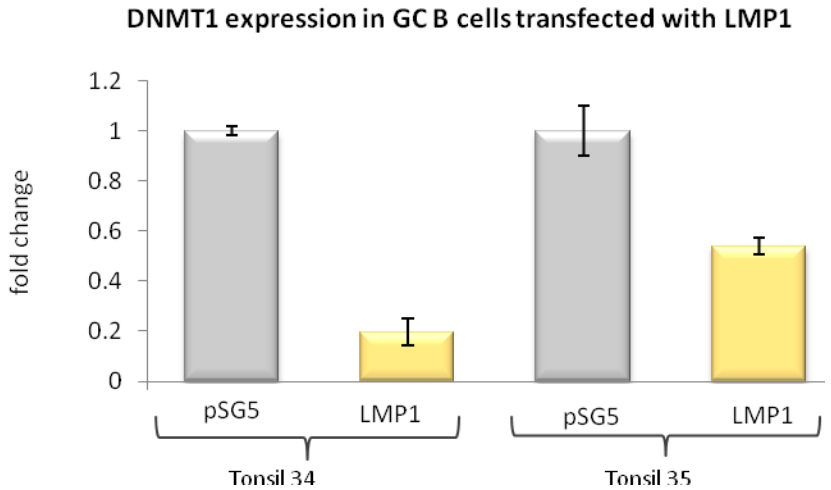



**Figure 4.6: Q RT-PCR of the relative quantity of DNMT1 in LMP1-expressing and non-expressing GC B cells.** Two different transfected tonsils were analysed in triplicate and the results presented as $2^{-\Delta\Delta CT}$ values compared with vector control (pSG5).

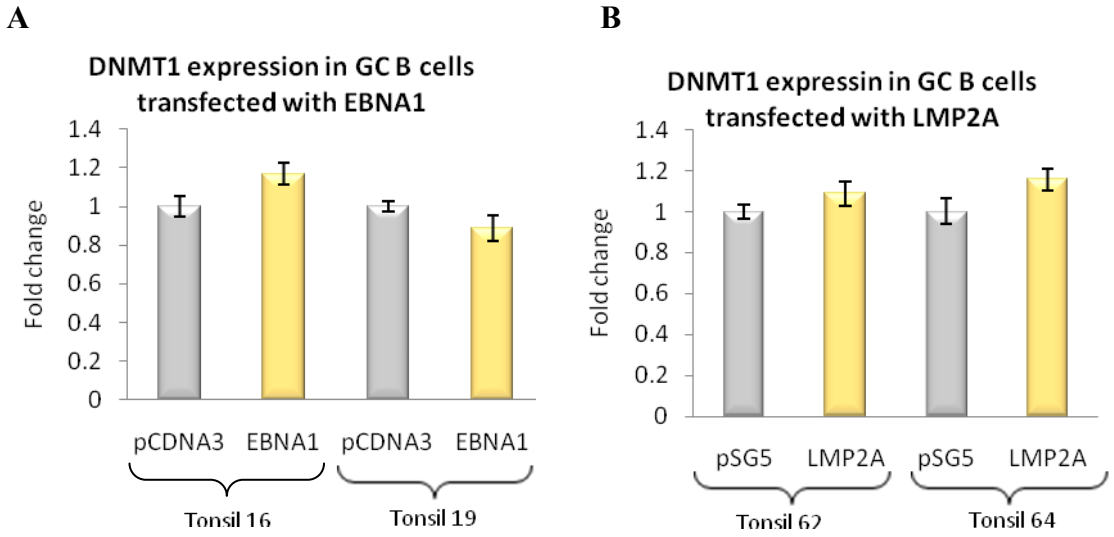



**Figure 4.7 Q RT-PCR of the relative quantity of DNMT1 in LMP2A or EBNA1-expressing and non-expressing GC B cells.** Different tonsils transfected with either LMP2A or EBNA1were analysed in triplicate and the results presented as $2^{-\Delta\Delta CT}$ values compared with vector controls. **A.** GC B cells transfected with EBNA1 **B.** GC B cells transfected with LMP2A.

## 4.6 DNMT1 is down-regulated in an LMP1-inducible system

As LMP1 was the only viral protein found to regulate DNMT1 in GC B cells, I next investigated if LMP1 could also down-regulate this enzyme in another B cell system. Towards this end I used a tetracycline-inducible expression system which is now described in brief. In this model, the EBV negative BL cell line DG75 was used to generate stable clones carrying a tetracycline-inducible LMP1 expression plasmid. These clones do not express LMP1 in the presence of tetracycline. However, once tetracycline is removed the LMP1 promoter is activated and its transcription is initiated (Floettmann *et al,*. 1996). When LMP1 is induced in this cell line it reaches levels similar to that observed in some LCLs. In Figure 4.8 A, I show DNMT1 expression in both LMP1 negative and positive clones, before and after the removal of tetracycline. In the LMP1 negative clones, the removal of tetracycline has little impact on the expression of DNMT1. In contrast, in the LMP1 positive clones, withdrawal of tetracycline is followed by a substantial decrease in DNMT1 expression. These results confirm in another system that LMP1 can down-regulate DNMT1. However western blotting performed on these samples revealed no change in the expression of DNMT1at the protein level (Figure 4.8 B).

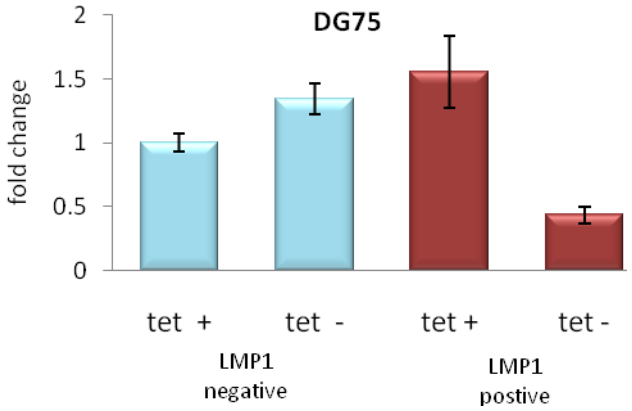



**Figure 4.8 A: Q RT-PCR analysis of DNMT1 expression in DG75 with tetracycline-inducible expression of LMP1.** Blue bars represent clones containing no LMP1 construct and red bars represent LMP1 containing clones. Tet + indicates cells grown in the presence of tetracycline and Tet- indicates cells with tetracycline removed. Assays were carried out in triplicate and results are presented as $2^{-\Delta\Delta CT}$ values compared with vector controls.

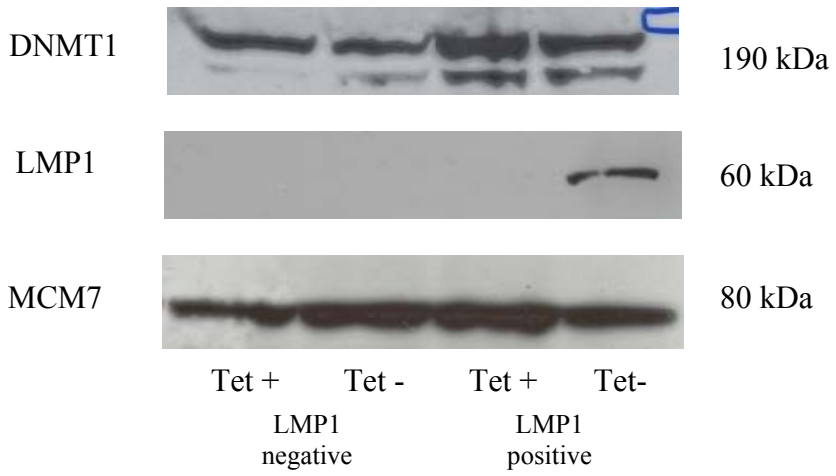



**Figure 4.8 B: Western blotting analysis of DNMT1 expression in DG75 with tetracycline-inducible expression of LMP1.** DNMT1 and LMP1 expression was detected using 30 µg protein. MCM7 served as a loading control.

## 4.7 Neither DNMT3A nor DNMT3B are regulated by EBV latent genes in GC B cells

I next investigated whether the expression of DNMT3A and DNMT3B in GC B cells was regulated by LMP1, LMP2A or EBNA1. For these experiments, I used RNA which had been obtained following transfection of GC B cells with these latent genes as described in section 4.6. Figure 4.9A shows that DNMT3A expression was decreased in one of the samples transfected with LMP2A but increased in the other. Transfection of GC B cells with LMP1 or EBNA1 did not cause any significant change in DNMT3A expression. (Figure 4.9 B and C). DNMT3B expression remained unchanged following transfection of GC B cells with LMP1, EBNA1 or LMP2A (Figure 4.10 A-C). The viral gene that regulates DNMT3A and DNMT3B remain to be identified.

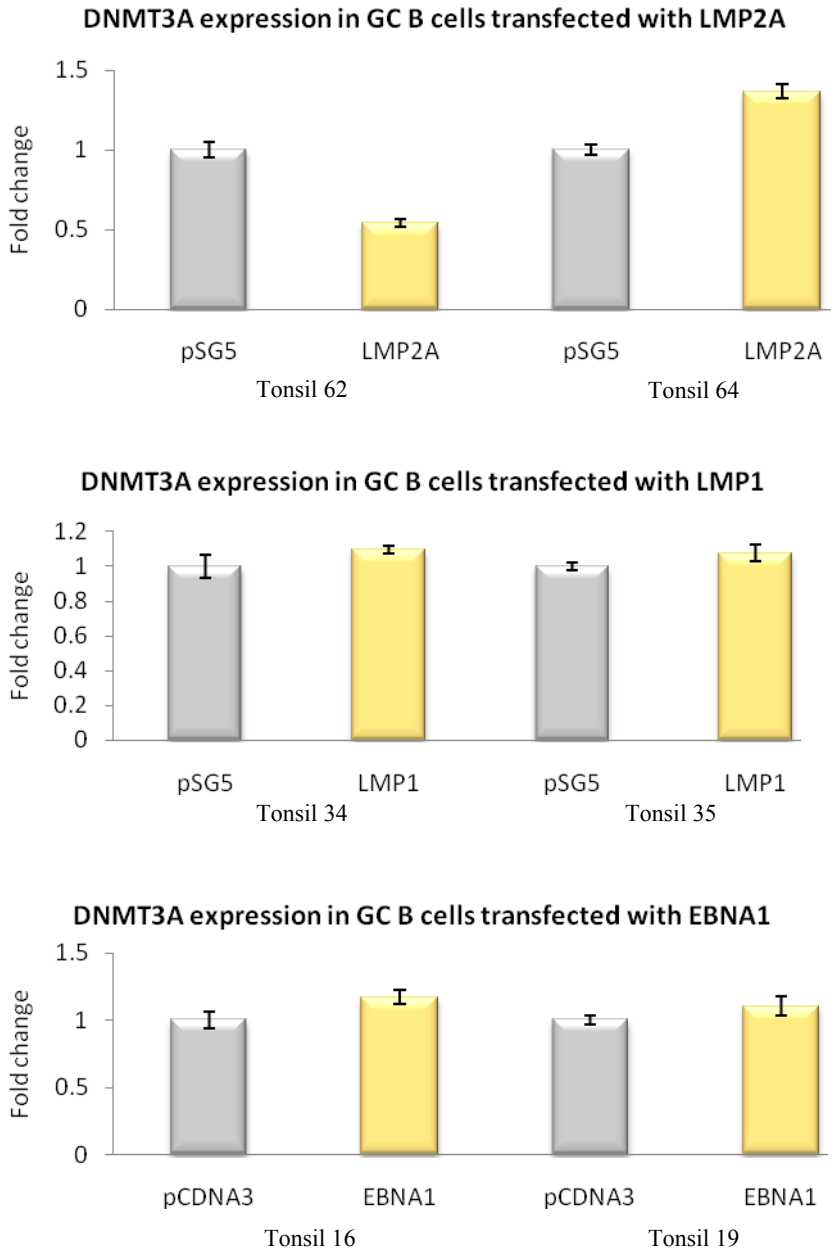



**Figure 4.9: DNMT3A expression in GC B cells transfected with LMP1, LMP2A or EBNA1 expression vectors**. Each assay was carried out in triplicate and the results shown are presented as $2^{-\Delta\Delta CT}$ values compared with suitable vector control. Standard error bars show the variation of DNMT3A expression across the three assays. **A.** Expression of DNMT3A in LMP2A transfected GC B cells. **B.** Expression of DNMT3A in LMP1 transfected GC B cells. **C.** Expression of DNMT3A in EBNA1 transfected GC B cells.

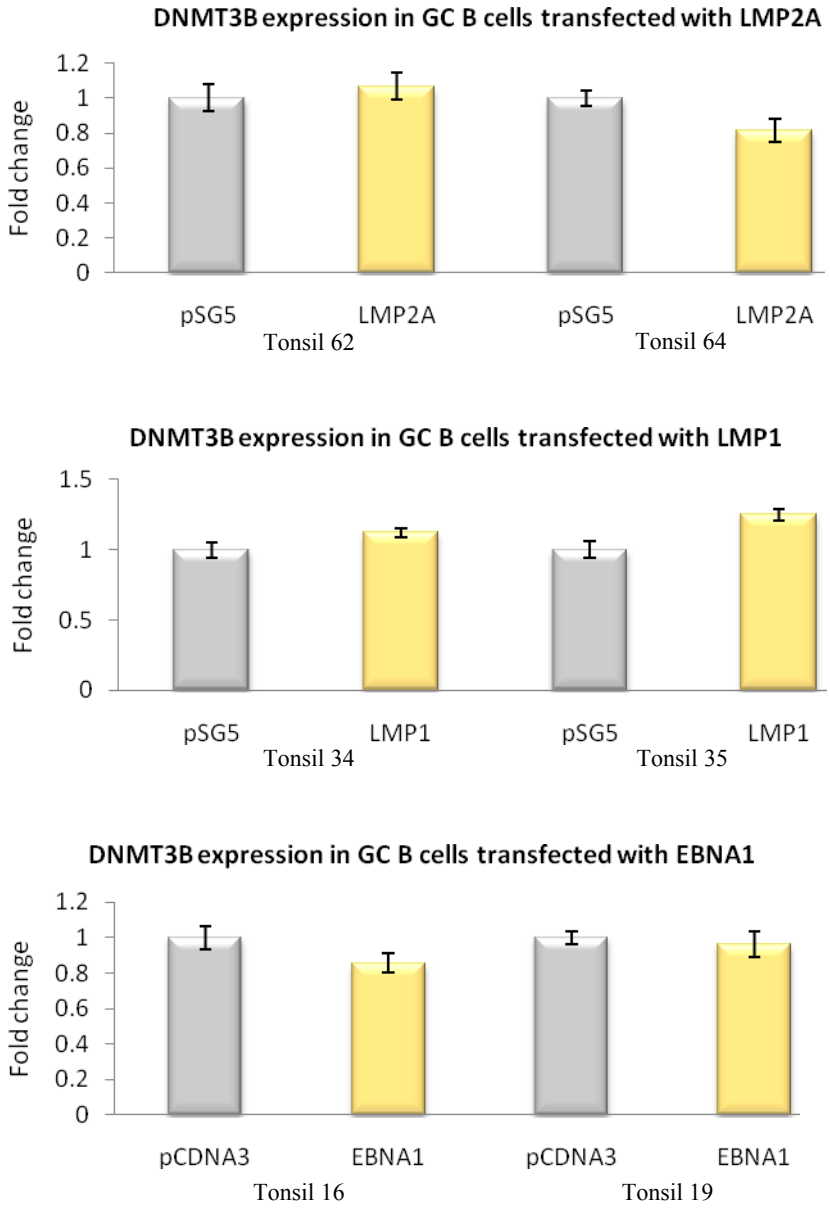



**Figure 4.10: DNMT3B expression in GC B cells transfected with LMP1, LMP2A or EBNA1 expression vectors**. Each assay was carried out in triplicate and the results shown are presented as $2^{-\Delta\Delta CT}$ values compared with suitable vector control. Standard error bars show the variation of DNMT3B expression across the three assays. **A**. Expression of DNMT3B in LMP2A transfected GC B cells. **B.** Expression of DNMT3B in LMP1 transfected GC B cells. **C.** Expression of DNMT3B in EBNA1 transfected GC B cells.

## Summary

In this section, I have shown that the expression of DNMT3A was increased, and that of DNMT1 and DNMT3B decreased soon after EBV infection. The expression of DNMT3A was increased, that of DNMT1 and DNMT3B was decreased in a panel of HL cell lines when compared with GC B cells. Finally, I have shown that DNMT1 is down-regulated in B cells by the EBV oncogene LMP1.

In the last section of this chapter, I show how the expression of the protein arginine methyltransferases, CARM1, PRMT1 and PRMT5, and the peptidylarginine deiminase, PADI4 changes following EBV infection of GC B cells. I profile the expression of these proteins across a panel of HL cell lines before investigating whether the expression of one or more of them is modulated by the EBV latent gene LMP1.

## 4.8 EBV infection of GC B cells modulates the expression of enzymes regulating arginine methylation

The microarray performed on GC B cells and LCLs six week post infection revealed up-regulation of the arginine methyltransferases, PRMT1, PRMT5 and CARM1 in LCLs compared to GC B cells, and down-regulation of the peptidyl-arginine deiminase, PADI4 (Table 4.2).

| NCBI gene symbol | Probe set ID | Fold change |
|---|---|---|
| CARM1 | 212512_s_at | 1.8 |
| PRMT1 | 206445_s_at | 1.76 |
| PRMT5 | 217786_at | 1.5 |
| PADI4 | 220001_at | -1.84 |

**Table 4.2: Microarray analysis results for the PRMTs.** SAM analysis was performed with a FDR of 5%.

I next validated changes in the expression of these enzymes using RNA harvested six weeks post infection from the three newly established GC B cell derived LCLs. Taqman assays used are listed in Section 2.4.4 (page 57).

Compared with uninfected GC B cells, Q-RT-PCR showed that the expression of PRMT1, PRMT5 and CARM1 was significantly up-regulated in all three LCLs (Figure 4.11 A, B and C). While, there was a decrease in PADI4 expression, the overall level of this transcript was very low in both the GC B cells and LCLs (Figure 4.11 D).

Western blotting confirmed that PRMT1, PRMT5 and CARM1 were up-regulated in LCLs as compared to GC B cells (Figure 4.12). The list of antibodies used is presented in section 2.5.3 (page 59). I did not perform a western blot for the PADI4 protein as there was no commercially available antibody at the time.

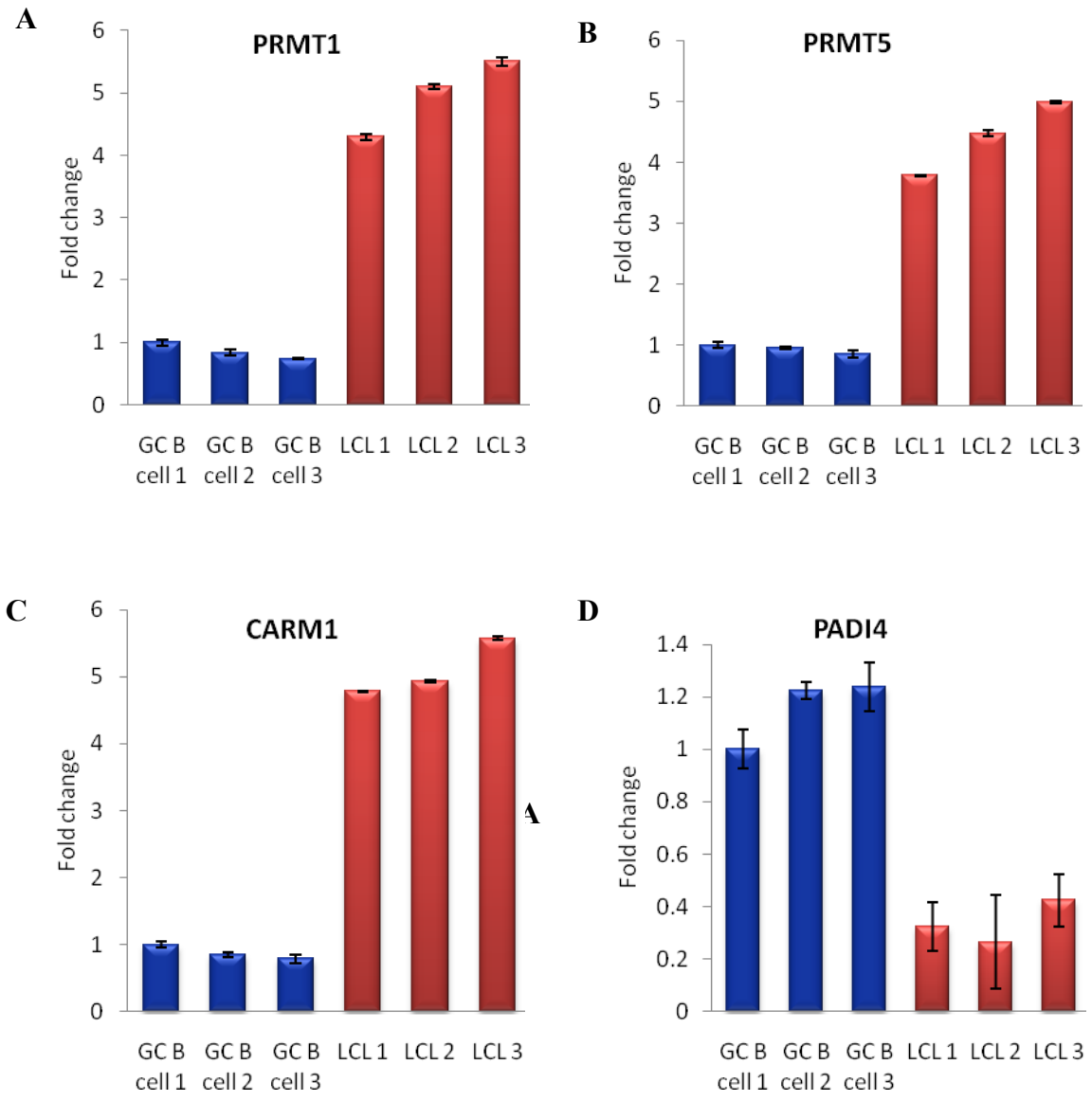



**Figure 4.11: PRMT and PADI4 mRNA expression in LCLs.** Blue bars represent three different patient GC B cell samples and red bars represent each of the LCLs. The GC B cells with the highest expression of the gene in question served as the reference with which other GC B cell samples and LCLs were compared. Each assay was carried out in triplicate and the results shown are presented as $2^{-\Delta\Delta CT}$ values. Standard error bars show the variation in relative PRMT expression across three independent experiments. **A.** PRMT1 expression **B**. PRMT5 expression. **C**. CARM1 expression and **D.** PADI4 expression.

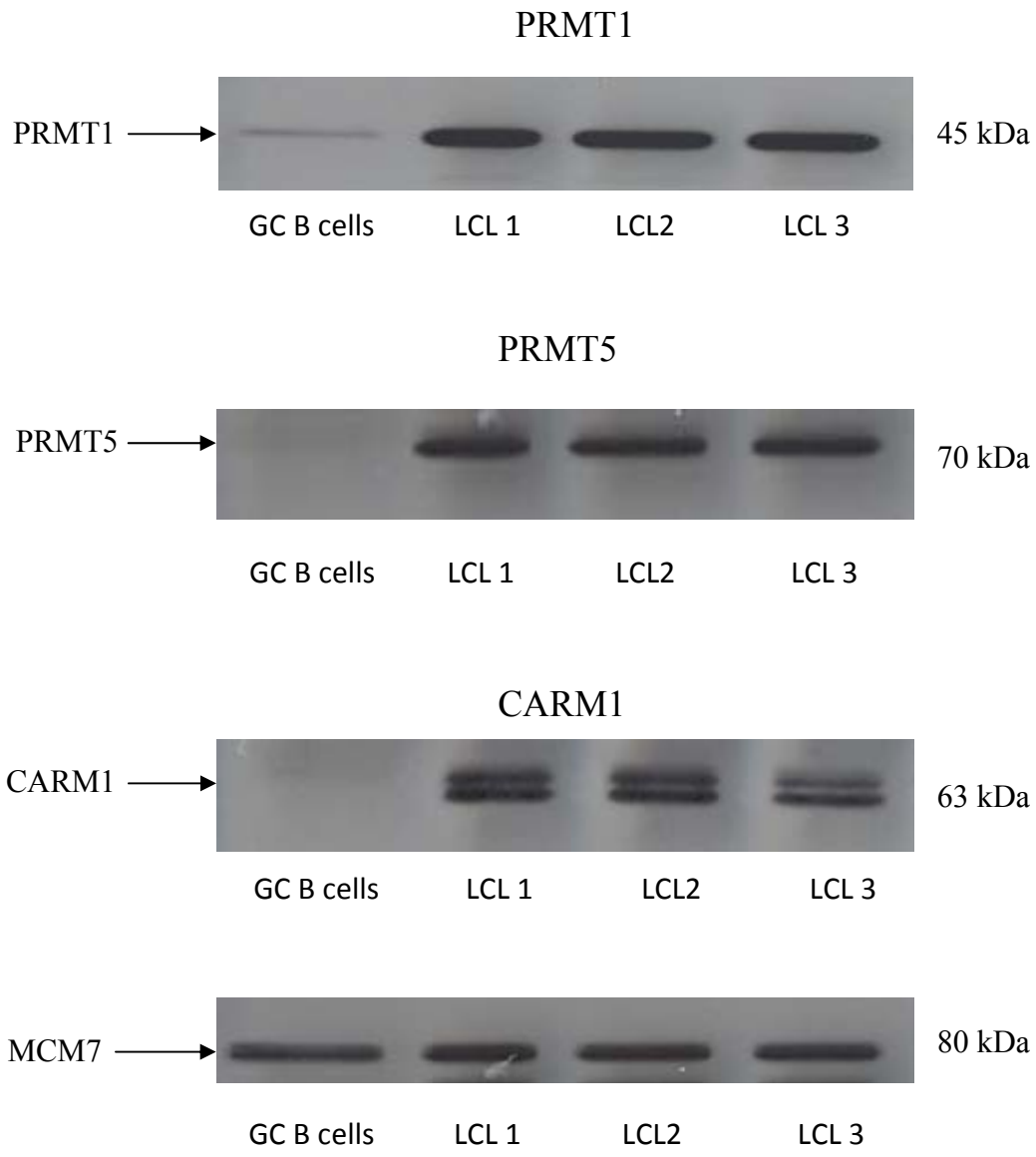



**Figure 4.12: Protein expression of PRMT1, PRMT5 and CARM1 in GC B cells and three GC B cell derived LCLs.** 20 ug of protein was used to detect each of the PRMTs. The same blot was used to measure protein levels for each PRMT and re-probed for MCM7 to ensure equal loading. The molecular weight for each protein is shown.

## 4.9 Enzymes regulating arginine methylation are differentially expressed in HL cell lines

I next explored the expression of the enzymes regulating arginine methylation in a panel of five HL cell lines. Compared with GC B cells, Q RT-PCR revealed the up-regulation of PRMT1 and PRMT5 in all HL cell lines (Figure 4.13 A and B). CARM1 was up-regulated in all cell lines with the exception of L450 in which it was down-regulated (Figure 4.13 C). PADI4 was down-regulated in all cell lines (Figure 4.13 D). Western blotting confirmed up-regulation of PRMT1, PRMT5 and CARM1 in all HL cell lines (Figure 4.14A-C).

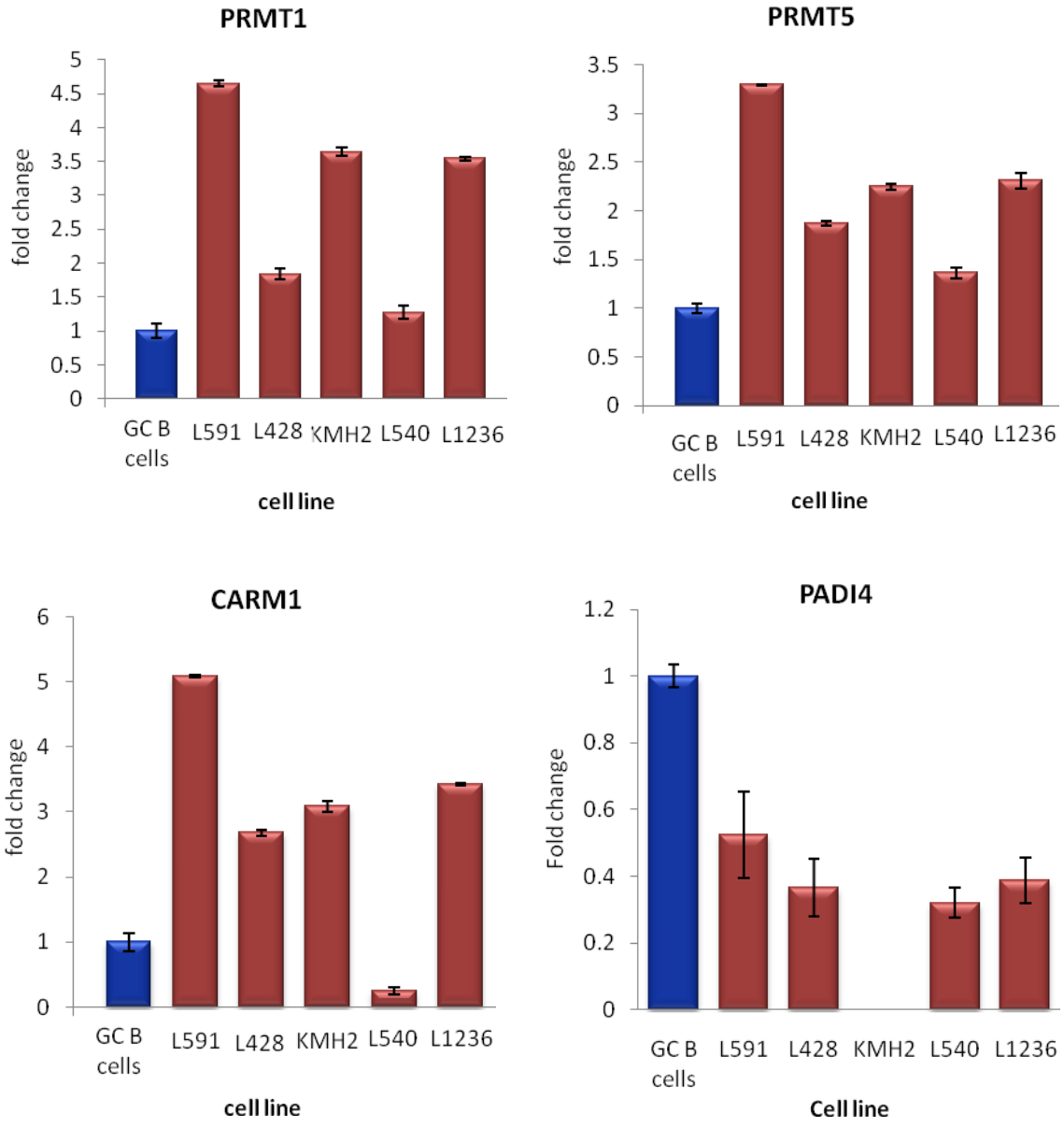



**Figure 4.13 A-D: Q RT-PCR analysis of PRMT and PADI4 transcripts expressed in a panel of HL cell lines.** The blue bar represents GC B cells and each of the red bars represents a different cell line. The expression of each cell line was compared to GC B cells. Each assay was carried out in triplicate and the results shown are presented as $2^{-\Delta\Delta CT}$ values. Standard error bars show the variation of PRMT expression across the three assays.

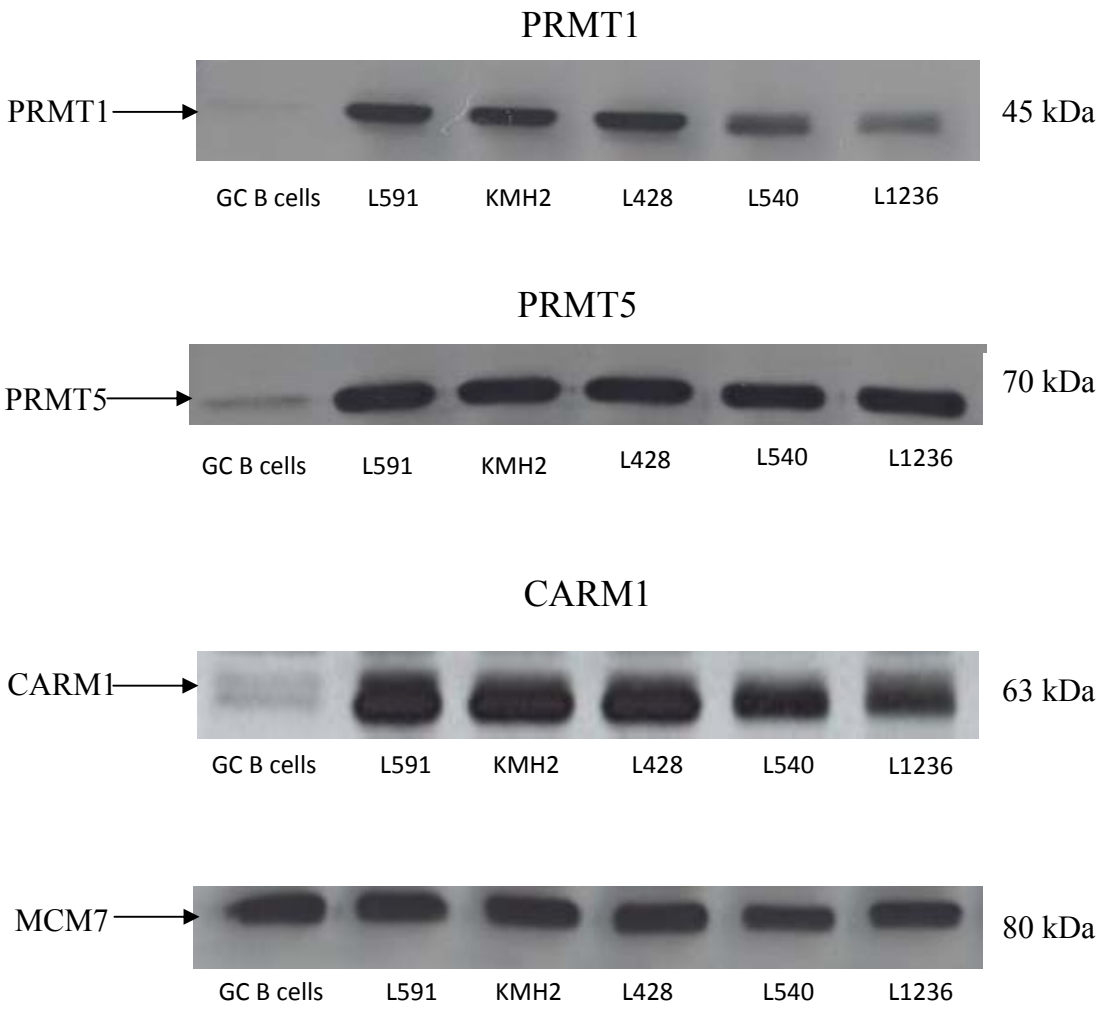



**Figure 4.14: Protein expression of PRMTs in a panel of HL cell lines.** 20ug of protein was used to detect protein expression of PRMT1, PRMT5 and CARM1. The same blot was used to measure protein levels for each PRMT and re-probed for MCM7 to ensure equal loading. The molecular weight of each protein is shown.

## 4.10 PRMT1 and PRMT5 are up-regulated by LMP1 in GC B cells

Having shown that EBV can up-regulate PRMT1, PRMT5 and CARM1 and down-regulate PADI4, I next explored whether the expression of these enzymes was regulated by LMP1 in B cells. For these experiments, I used RNA from GC B cells obtained from different patients transfected with either a pSG5-LMP1expression vector or with a pSG5 vector control (Vockerodt *et al*., 2008). I was able to show using Q-RT PCR and RNA isolated 24 hours post transfection that LMP1 up-regulated PRMT1 in both GC B cell samples (Figure 4.15 A). PRMT5 was up-regulated in only one of the samples (Figure 4.15 B) and neither CARM1 nor PADI4 were significantly changed (Figure 4.15 C and D).

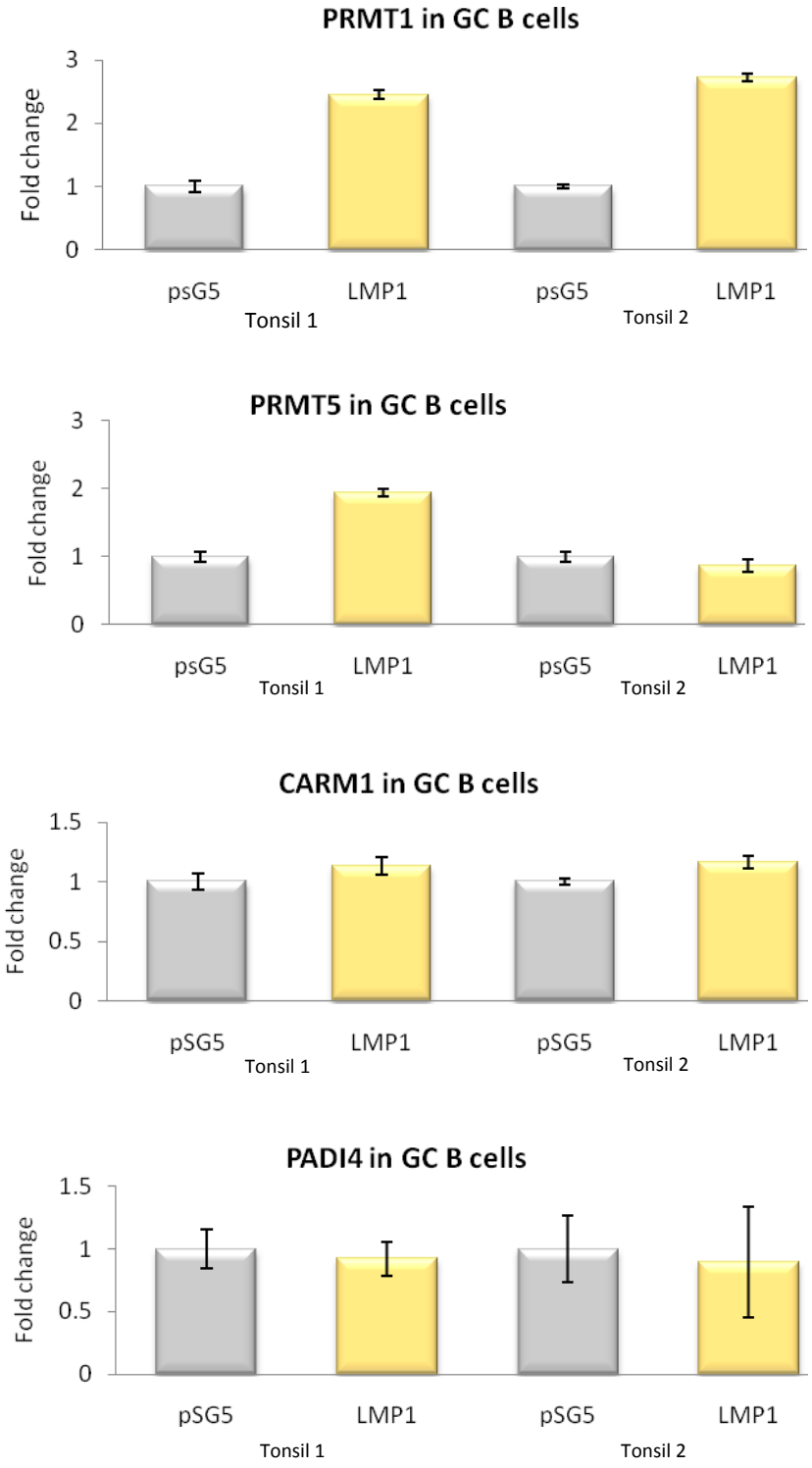
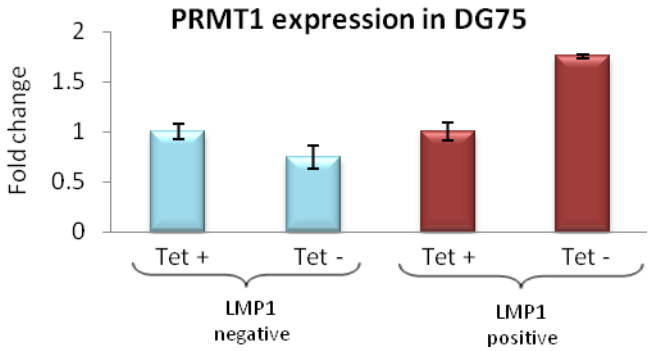



**Figure 4.15: Q RT-PCR of the relative quantity of PRMTs and PADI4 in LMP1-expressing and non-expressing GC B cells.** All samples were analysed in triplicate and are presented as $2^{-\Delta\Delta CT}$ values compared with vector control (pSG5). Standard error bars show the variation of expression across the three assays.

## 4.11 PRMT1 and PRMT5 are up-regulated in an LMP1-inducible system

Using the tetracycline-inducible expression system described earlier (section 4.6) I was able to show using Q RT-PCR that LMP1 up-regulated both PRMT1 and PRMT5 in the DG75 cell line (Figure 4.16).

**A**




**B**

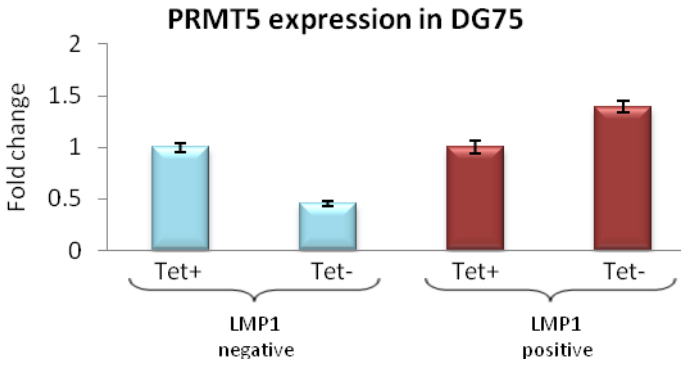



**Figure 4.16: Q RT-PCR analysis of PRMT1 and PRMT5 transcripts in cell lines with tetracycline-inducible expression of LMP1.** Blue bars represent clones containing no LMP1 construct and red bars represent LMP1 containing clones. Tet + indicates cells grown in the presence of tetracycline and Tet- indicates cells that have had tetracycline removed. Assays were carried out in triplicate and results are presented as $2^{-\Delta\Delta CT}$ values compared with vector controls. **A.** PRMT1 expression. **B**. PRMT5 expression.

**Summary**

In this section, I have shown that the expression of the arginine methyltransferases, CARM1, PRMT1 and PRMT5, is up-regulated, and that of the peptidylarginine deiminase, PADI4, down-regulated following EBV infection of GC B cells. Whereas the expression of CARM1, PRMT1 and PRMT5 is increased, that of PADI4 is decreased in HL cell lines when compared with GC B cells. Finally, I have shown that PRMT1 and PRMT5 are up-regulated in B cells by the EBV oncogene LMP1.

Chapter 5

# The impact of EBV induced changes in the expression of DNA methyltransferases on the methylation and expression of viral genes

**Aims of Chapter 5**

1. To determine the DNA methylation status of the Wp and Cp promoter in GC B derived LCLs
2. To measure the binding of the DNA methyltransferases to the Wp promoter in
   a) GC B derived LCLs
   b) a latency I Burkitt's lymphoma cell line
   c) a "Wp only" LCL
   d) an LCL containing 11 *Bam*HI W copies
   e) an LCL containing 2 *Bam*HI W copies
3. To determine the impact on viral gene expression, of DNMT1 and DNMT3B over-expression in GC B cell derived LCLs.

In the first section of this chapter, I investigate the impact of EBV-induced changes in the expression of the DNMT on the viral life cycle. I first show that the Wp promoter is methylated in GC B cell derived LCLs before investigating how DNMT binding to this promoter varies in EBV positive cell lines with different latency patterns and with different numbers of Wp repeats.

## 5.1 The Wp promoter is methylated in GC B cell derived LCL

The EBV promoter, Wp is known to become methylated soon after primary B cell infection (Tierney *et al.,* 2000). To determine whether Wp was methylated in GC B cell derived LCLs, I designed using the methods described in section 2.10.2 ,two sets of pyrosequencing primers which encompass the promoter regions around the YY1 binding sites (region 2) and BSAP binding site (region 1, Figure 5.1). Pyrosequencing was carried out twice on each of the LCL samples ( primer sequences are listed in page 78). In LCL 1, the average methylation across 5 CpGs was 29% in region 1 of Wp, and 60% in region 2 (Figure 5.2). Methylation levels were consistent across these three LCLs (Table 5.1).

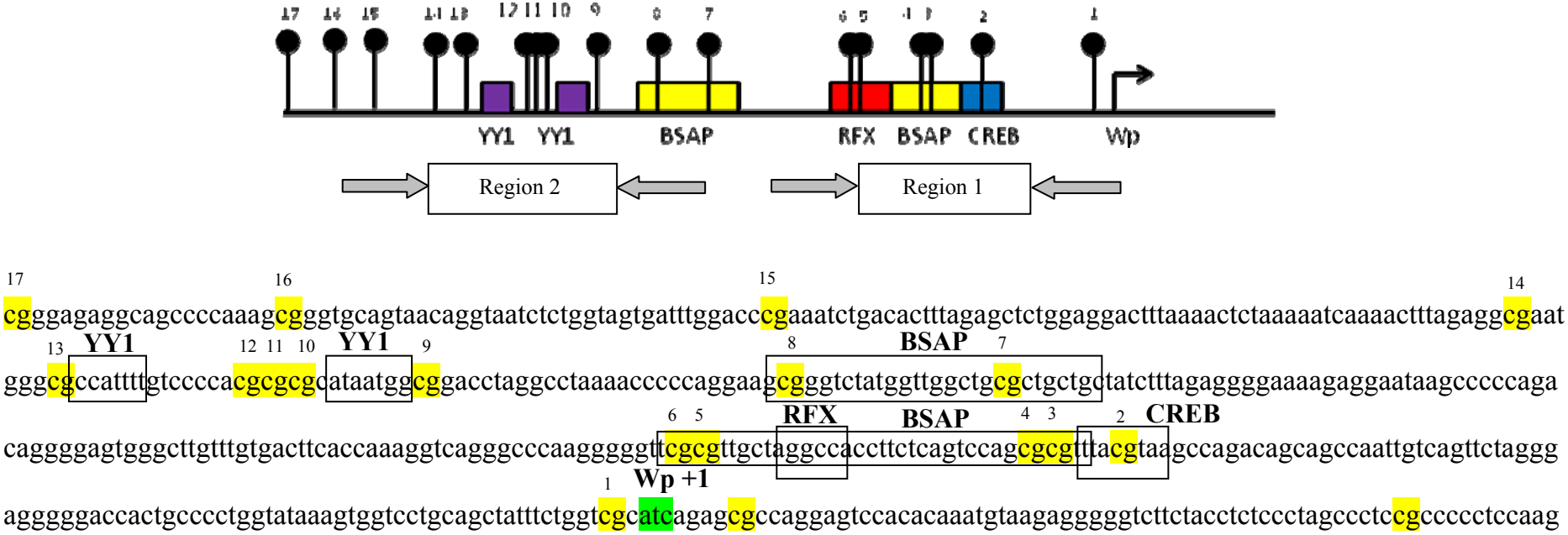



**Figure 5.1. DNA sequence of the EBV *Bam*HI Wp promoter regulatory region**. Coloured boxes indicate the binding sites for the transcription factors YY1, BSAP, RFX and CREB. The black circles labelled 1 to 15 denote the CpG dinucleotides within the Wp promoter. The arrows indicate the two regions analysed by pyrosequencing. Within the sequence the CpGs are highlighted in yellow and numbered corresponding to their position on the diagram above. The transcription start site is highlighted in green.

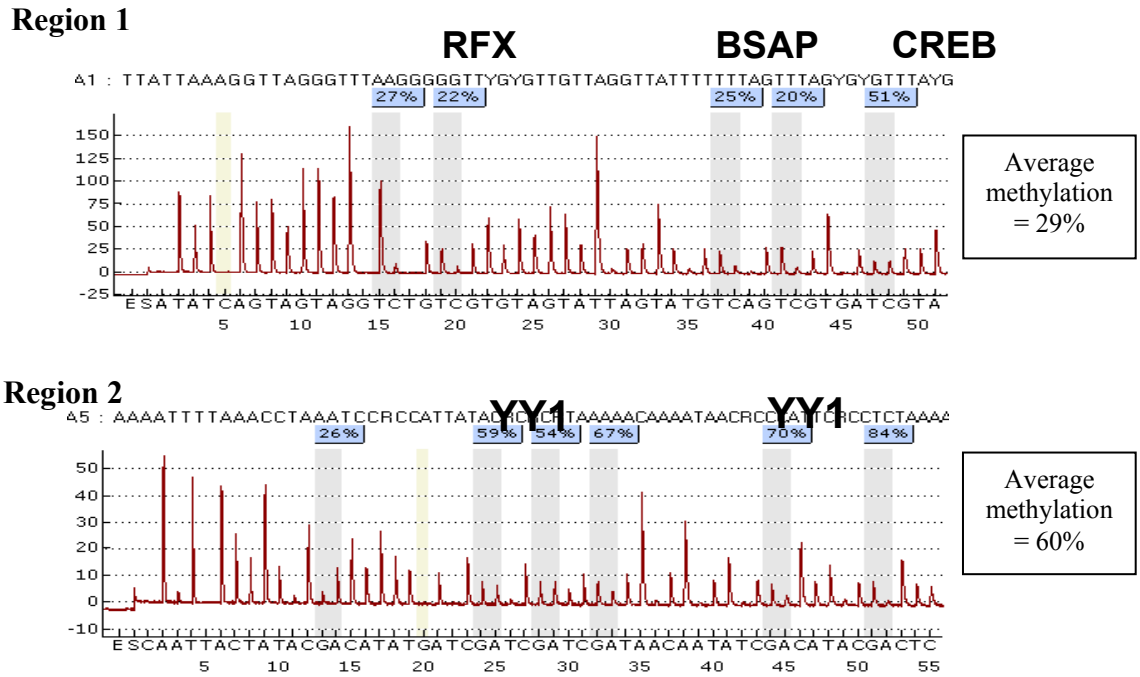



**Figure 5.2: Pyrogram results showing the percentage methylation at each of the CpG dinucleotides for two regions analysed within the Wp promoter.** The binding sites for transcription factors known to bind within that region are labelled above the pyrogram in bold. The average methylation across each region is reported in the box on the right.

| | CpG 1 | CpG 2 | CpG 3 | CpG 4 | CpG 5 | CpG 6 | CpG 7 | CpG 8 | CpG 9 | CpG 10 | CpG 11 | **Average methylation** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **LCL1** | 27 | 22 | 25 | 20 | 51 | 26 | 59 | 54 | 67 | 70 | 84 | 50.5% |
| **LCL 2** | 32 | 23 | 29 | 19 | 47 | 25 | 63 | 54 | 69 | 77 | 80 | 51.8% |
| **LCL 3** | 35 | 21 | 32 | 25 | 53 | 27 | 70 | 55 | 72 | 75 | 93 | 55.8% |

**Table 5.1: Summary of methylation results at each CpG in LCL1, LCL 2 and LCL 3**. The average methylation over the regions is reported in the box on the far right hand side.

## 5.2 Optimization of X-ChIP antibodies

To test the specificity and sensitivity of each of the DNMT antibodies to be used in the cross-linked chromatin immunoprecipitation (X-ChIP), a series of preliminary experiments were performed using a protocol optimized by Dr. Laura O'Neill (Division of Immunity and Infection, University of Birmingham). Two different antibodies against DNMT1, a polyclonal antibody (Abcam) and a monoclonal antibody (Imgenex), were compared by assessing their ability to bind to the HoxA7 promoter in HeLa cells. DNMT1 has already been shown to bind to "region E" but not to "region F" of this promoter in these cells (Wu *et al*., 2008). A polyclonal antibody to DNMT3A (Imgenex) previously shown to be successful in X-ChIP experiments, and a ChIP-grade monoclonal antibody to DNMT3B (Abcam) were tested for binding to the GAPDH and Sat 2 promoters in an LCL (Leu *et al.,* 2004). Both DNMT3A and DNMT3B have previously been shown to bind to Sat 2 but not to GAPDH in an LCL (Geiman *et al.,* 2004). Details of the antibodies and primers used in these experiments are provided in section 2.13.3 Page 84) IPs were performed as described in section 2.13 and the DNA purified from IP and input samples analysed using SYBR green Q-PCR. The percent input method was used to calculate the enrichment in the target sequences as described in section 2.13.6.

Q-PCR showed that region E but not region F of the HoxA7 promoter was significantly enriched in both DNMT1 IPs compared to the control IgG and no antibody IP, and these findings are consistent with the X-ChIP results obtained by Wu *et al.,* 2008 (Figure 5.3 A and B). To assist with the interpretation of these figures I should point out that whereas a substantial increase in the PCR product in the DNMT IP compared to IgG and no antibody controls indicates evidence of specific binding of that DNMT to that region, no enrichment in the DNMT IP suggests that there has been no binding. As a greater enrichment was detected

using the Abcam antibody this was used in subsequent X-ChIP experiments. I next assessed the DNMT3A and DNMT3B antibodies in a LCL. When compared to controls a significant enrichment of DNMT3A and DNMT3B was detectable at the Sat 2 promoter (Figure 5.3 C). As expected the GAPDH promoter was found to be negative for DNMT binding (Figure 5.3 D).

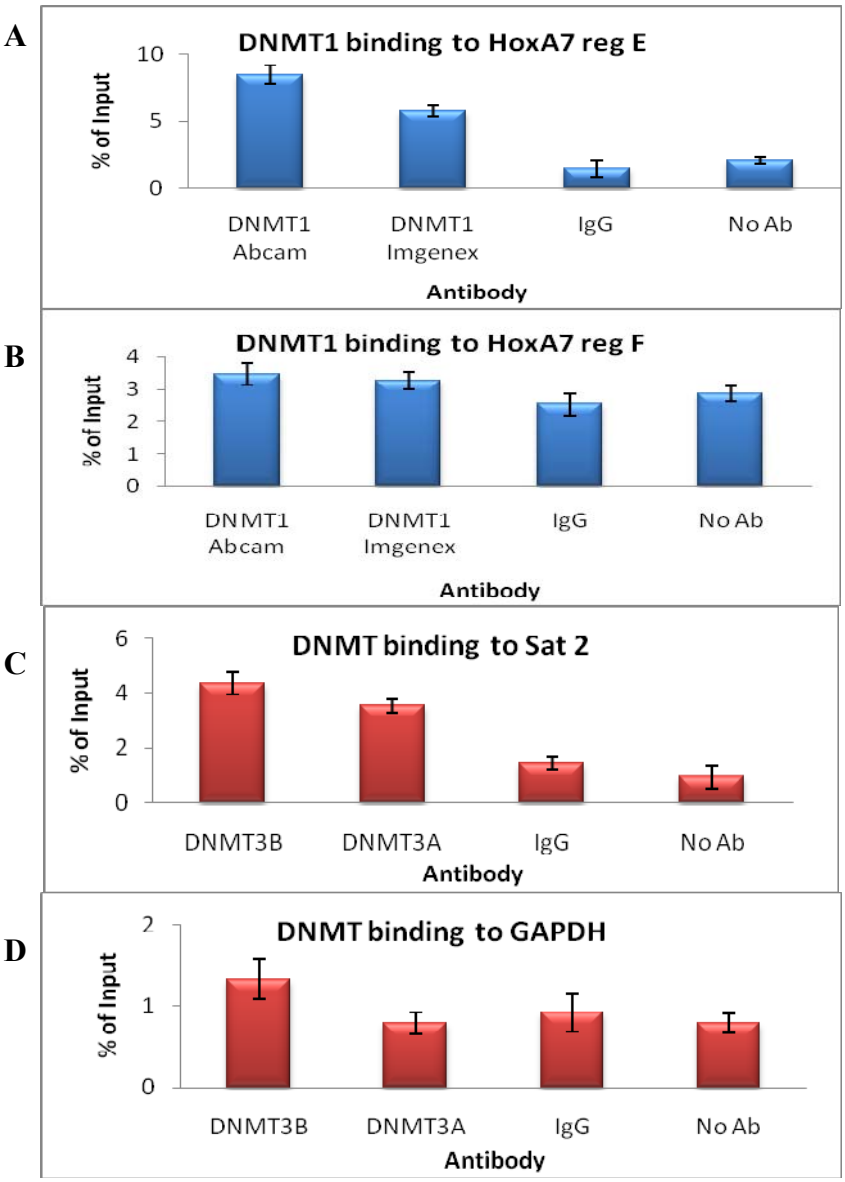



**Figure 5.3: Q-PCR analysis testing the efficiency of DNMT antibodies used in IP.** HeLa cells were used to test the efficiency of two DNMT1 antibodies. LCLs were used to test DNMT3A and DNTM3B antibodies. All results are reported as a % of input and compared to IgG and no antibody controls. **A and B.** DNTM1 binding to the HoxA7 promoter. Region E served as a positive control and region F as a negative. **C and D.** DNMT3A and DNMT3B binding to the highly methylated Sat 2 region and unmethylated GAPDH promoter.

## 5.3 DNMT3A binds to the methylated Wp promoter in GC B cell derived LCLs

Having successfully established the X-ChIP technique with each of the DNMT antibodies, I next studied the binding of DNMT1, DNTM3A and DNMT3B to Wp in the GC B cell derived LCLs. Figure 5.4 shows where the overlapping PCR primers were designed for 9 regions of the Wp promoter, and one for the Cp promoter, which is known to be unmethylated in LCLs and therefore provides a negative control (Tierney *et al.,* 2000). The full list of primer sequences is shown in section 2.13.6, page 85. X-ChIP experiments were performed as before.

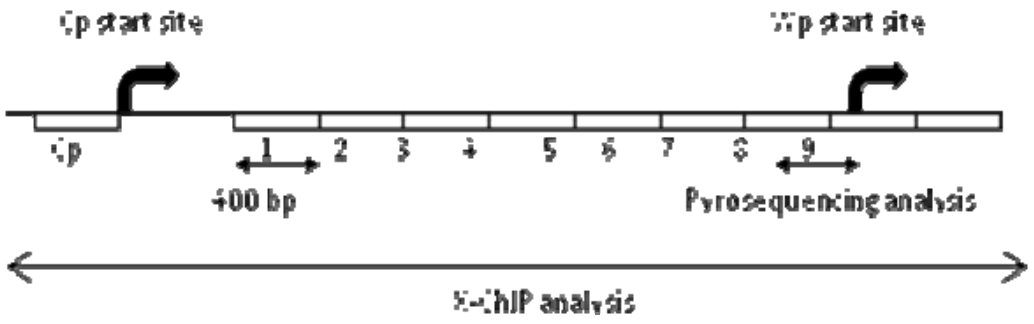



**Figure 5.4 Regions of Wp covered by X-ChIP**. The nine sets of primers covering Wp and one at Cp are shown. Each primer was between 300-400 bp in length. The region covered by pyrosequencing is also shown.

### 5.3.1 DNMT1 does not bind to the Wp promoter

In each of the three LCLs studied, DNMT1 binding to regions within the Wp or Cp promoter did not exceed that observed in the IgG and in the no antibody controls (Figure 5.5).

### 5.3.2 DNMT3B does not bind to the Wp promoter

In each of the three LCLs studied, DNMT3B binding to regions within the Wp or Cp promoter did not exceed that observed in the IgG and in the no antibody controls (Figure 5.6). Sat 2 (positive control) but not GAPDH (negative control) was found to bind DNMT3B.

### 5.3.3 DNMT3A binds to the Wp promoter

In each of the three LCLs studied, the binding of DNMT3A to regions within Wp was significantly greater than that observed in the IgG and in the no antibody controls (Figure 5.7). Enrichment was greatest in region 6 but was also detected in region 7. Sat 2 (positive control) but not GAPDH (negative control) was found to bind DNMT3A.

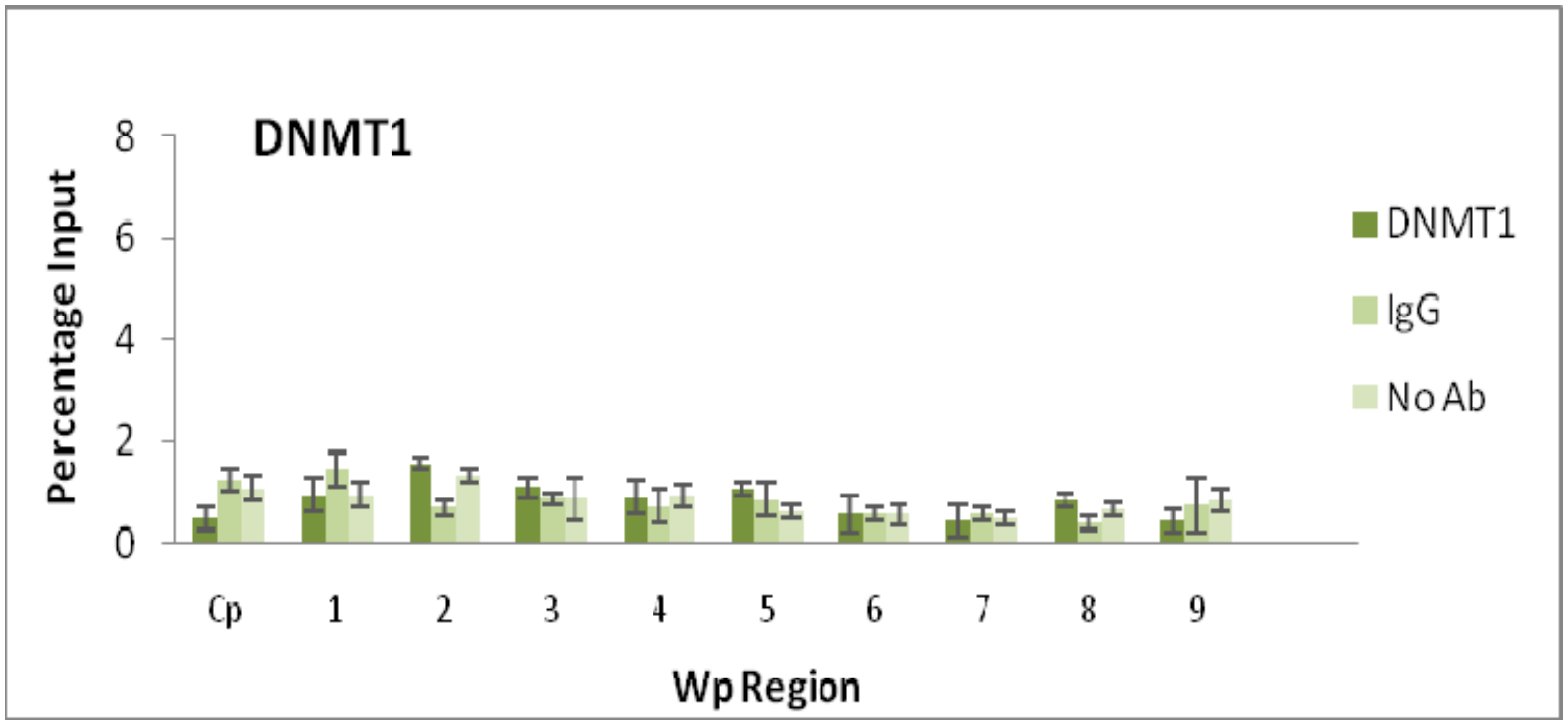



**Figure 5.5: SYBR green Q-PCR results of DNMT1 binding to the Wp promoter using X-ChIP.** DNMT1 binding to nine regions of the Wp and Cp promoter was assessed using X-ChIP. Binding was reported as a % of input. IgG and no antibody IPs were used as controls to which the antibody of interest was compared to. Each assay was carried out in duplicate and standard error bars show the variation across the assays.

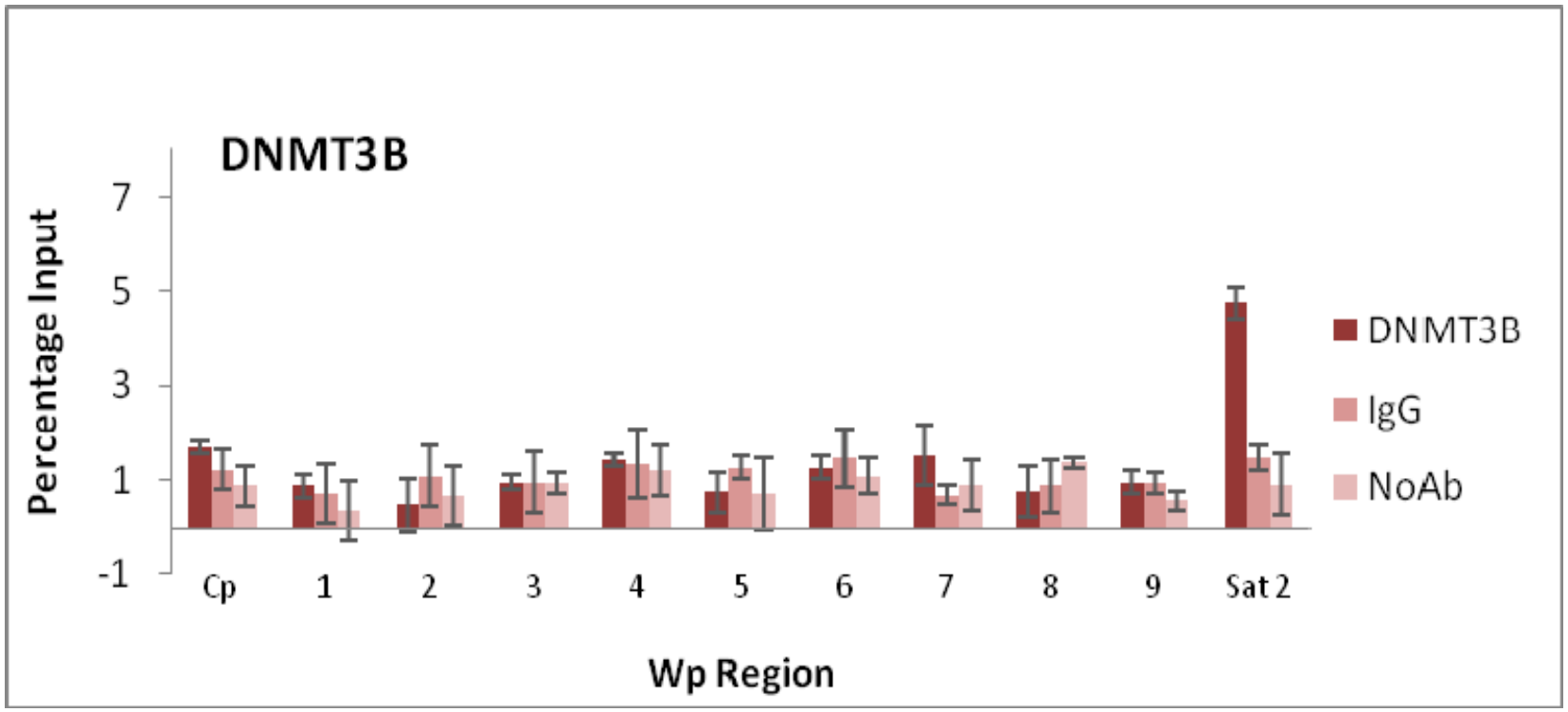



**Figure 5.6: SYBR green Q-PCR results of DNMT3B binding to the Wp promoter using X-ChIP.** DNMT3B binding to nine regions of the Wp and Cp promoter was assessed using X-ChIP. Methylated Sat 2 was used as a control for binding. Binding was reported as a % of input. IgG and no antibody IPs were used as controls to which the antibody of interest was compared to. Each assay was carried out in duplicate and standard error bars show the variation across the assays.

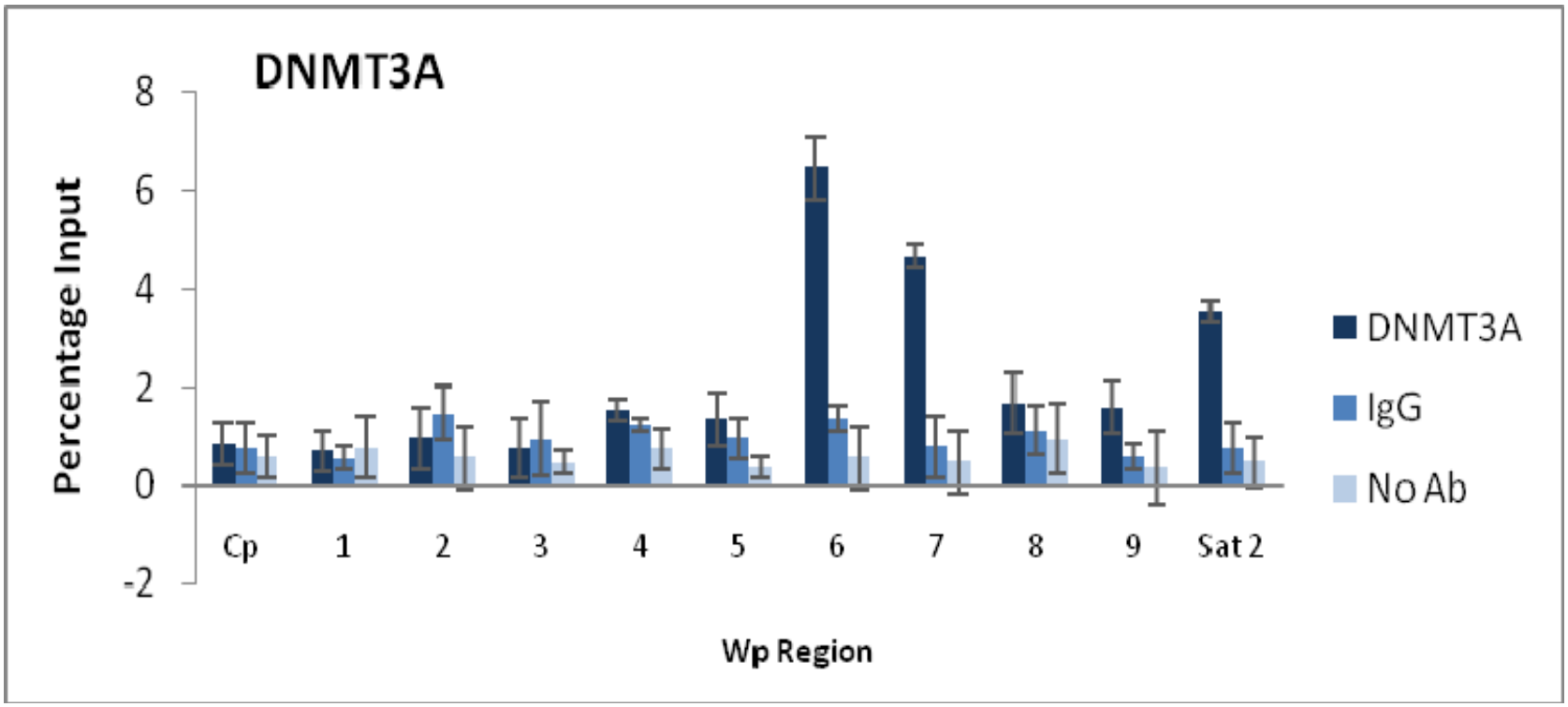



**Figure 5.7: SYBR green Q-PCR results of DNMT3A binding to the Wp promoter using X-ChIP.** DNMT3A binding to nine regions of the Wp and Cp promoter was assessed using X-ChIP. Methylated Sat 2 was used as a control for binding. Binding was reported as a % of input. IgG and no antibody IPs were set up as controls to which the antibody of interest was compared to. Each assay was carried out in duplicate and standard error bars show the variation across the assays.

So far I have studied only GC B cell derived LCLs infected with EBV carrying 11 copies of Wp. In the next section I explore DNMT binding to Wp in other EBV positive cell lines with different latency patterns and also different numbers of Wp repeats. All nine regions of Wp were analysed, however only regions 5, 6 and 7 are shown as these were the only regions that DNMT binding was detected.

## 5.4 Minimal binding of DNMT to the unmethylated Wp promoter in a "Wp only" LCL

X50-7 is a Wp using LCL, in which Wp is hypomethylated and transcriptionally active (Hutchings *et al.,* 2006). In this cell line, there was a minute excess of DNMT3A and DNMT3B binding to regions 5 and 6 in the Wp promoter as compared to that observed in the IgG and in the no antibody controls (Figure 5.8). Although initially thought to be background noise this minimal binding was a consistent finding across five different experiments.

## 5.5 DNMT3A and DNMT3B binds the methylated Wp promoter in a latency I Burkitt's Lymphoma cell line

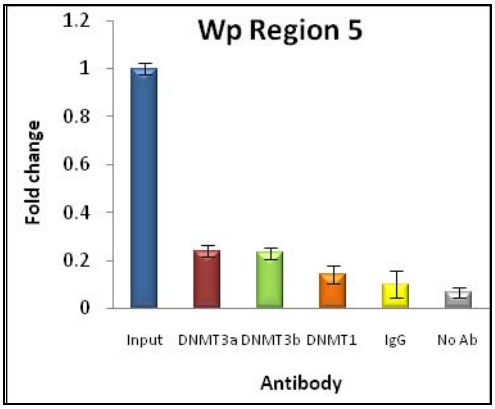
Rael-BL, is a standard latency I Burkitt's lymphoma cell line in which Wp is reported to be silenced and methylated (Hutchings *et al.,* 2006). In this cell line, the binding of DNMT3A and DNMT3B to regions 5 and 6, and to a lesser extent, region 7 within the Wp promoter was significantly greater than that observed in the IgG and in the no antibody controls (Figure 5.9). The binding of DNMT1 to regions 5 and 7 within the Wp promoter was also significantly greater than that observed in the IgG and in the no antibody controls. There was no evidence of DNMT binding to the Cp promoter.

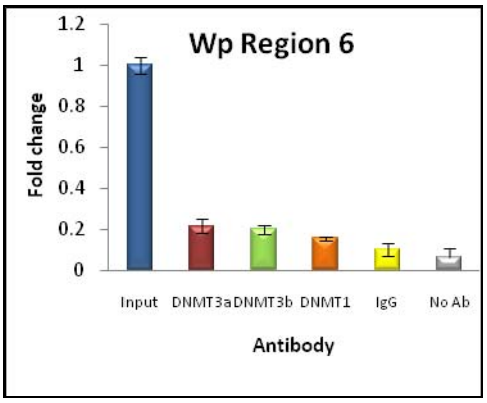
## 5.6 DNMT3A and DNMT3B bind the methylated Wp promoter in the 11W EBV LCL

The 11W LCL was constructed by Dr. Rose Tierney (School of Cancer Sciences, University of Birmingham) using a recombinant virus containing eleven *Bam*HI W repeats. Wp is reported to be methylated and silenced in this LCL (Hutchings *et al.,* 2006). In this cell line, the binding of DNMT3A in regions 5, 6 and 7 within the Wp promoter was significantly greater than that observed in the IgG and in the no antibody controls (Figure 5.10). Binding of DNMT3B and DNMT1 was occasionally detected but was not a consistent finding. No enrichment was detected in the Cp promoter.

## 5.7 DNMT3A and DNMT3B bind the "unmethylated" Wp promoter in the 2W EBV LCL

The 2W LCL was constructed by Dr. Rose Tierney (School of Cancer Sciences, University of Birmingham) using a recombinant virus containing two *Bam*HI W repeats. Wp is reported to be unmethylated and transcriptionally active in this LCL (Hutchings *et al.,* 2006). In this cell line, the binding of DNMT3A in region 6 within the Wp promoter was significantly greater than that observed in the IgG and in the no antibody controls (Figure 5.11). No consistent binding of DNMT3B and DNMT1 could be detected. No enrichment was detected in the Cp promoter.



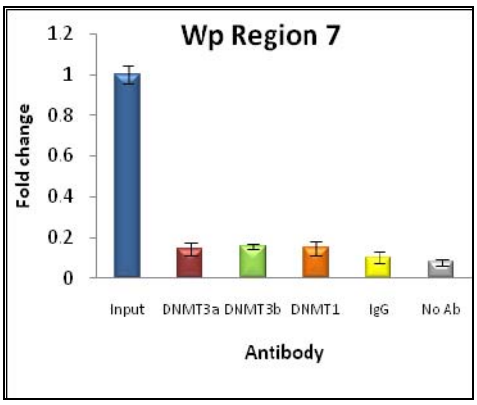




**Figure 5.8: X-ChIP results showing DNMT3A, DNMT3B and DNMT1 binding to Wp in the X50-7 LCL.** IPs were set up using a DNMT antibody or using IgG or no antibody as controls. DNA was purified from IPs and the input sample and assayed in triplicate using SYBR green Q-PCR for regions in the Wp promoter; 5, 6 and 7 . Results are expressed as a percentage of input. Results are representative of experiments performed five times.

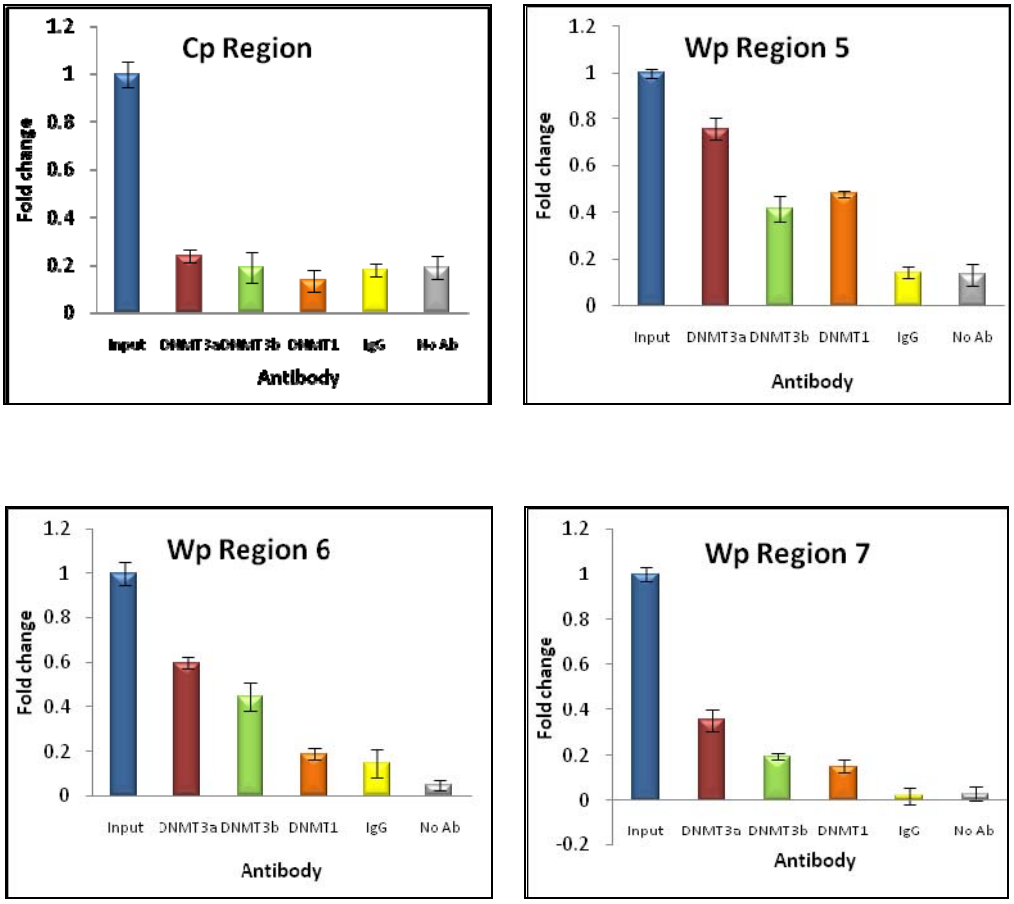



**Figure 5.9: X-ChIP results showing DNMT3A, DNMT3B and DNMT1 binding to Wp in Rael - BL.** IPs were set up using a DNMT antibody or using IgG or no antibody as controls. DNA was purified from IPs and the input sample and assayed in triplicate using SYBR green Q-PCR for regions in the Wp promoter; 5, 6, 7 and also for the Cp promoter (negative control). Results are expressed as a percentage of input. Results are representative of an experiment performed three times.

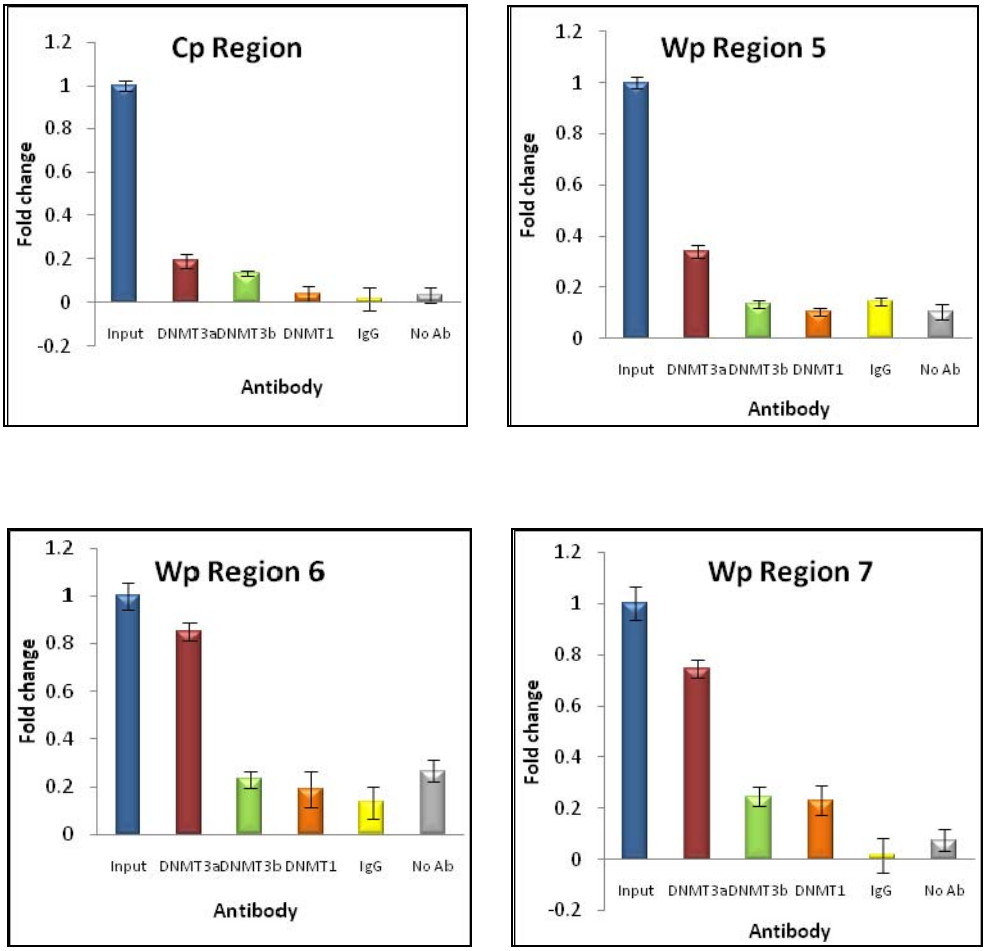



**Figure 5.10: X-ChIP results showing DNMT3A, DNMT3B and DNMT1 binding to Wp in the 11W EBV LCL.** IPs were set up using a DNMT antibody or using IgG or no antibody as controls. DNA was purified from IPs and the input sample and assayed in triplicate using SYBR green Q-PCR for regions in the Wp promoter; 5, 6, 7 and also for the Cp promoter (negative control). Results are expressed as a percentage of input. Results are representative of an experiment performed three times.

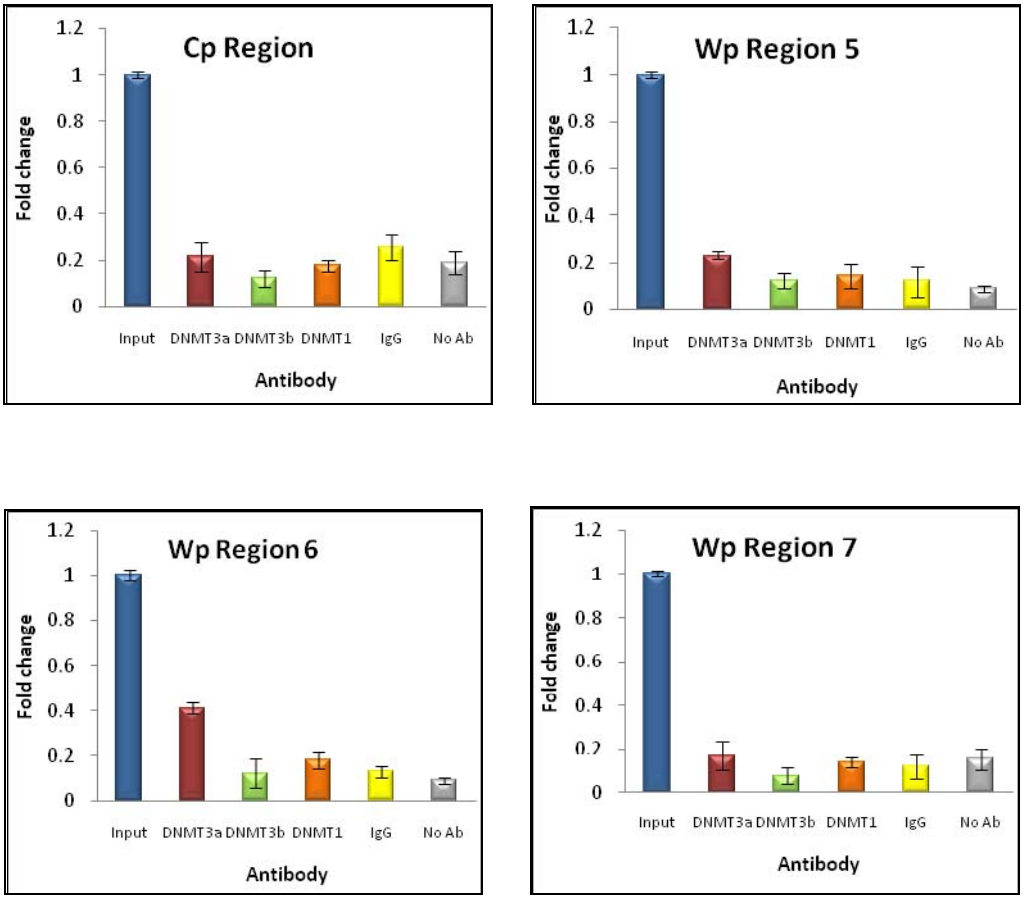



**Figure 5.11: X-ChIP results showing DNMT3A, DNMT3B and DNMT1 binding to Wp in the 2W EBV LCL.** IPs were set up using a DNMT antibody or using IgG or no antibody as controls. DNA was purified from IPs and the input sample and assayed in triplicate using SYBR green Q-PCR for regions in the Wp promoter; 5, 6, 7 and also for the Cp promoter (negative control). Results are expressed as a percentage of input. Results are representative of an experiment performed three times.

## 5.8 Methylation status of the Wp promoter in the 2W EBV LCL-revisited

In the preceding section I have shown that DNMT3A binds to the Wp promoter in the 2W EBV LCL. This was a surprising finding because the Wp promoter in this cell line is reported to be unmethylated (Hutchings *et al.,* 2006). Therefore, using pyrosequencing, I examined the methylation status of this promoter and found that methylated CpGs could be detected in Wp (Figure 5.12). On revisiting the data presented in the manuscript, I found that some clones were methylated at certain CpGs when examined using bisulphite genomic sequencing Therefore the results of my X-ChIP experiment demonstrating DNMT3A binding to the Wp promoter are consistent with the results of my pyrosequencing and with the data originally presented.

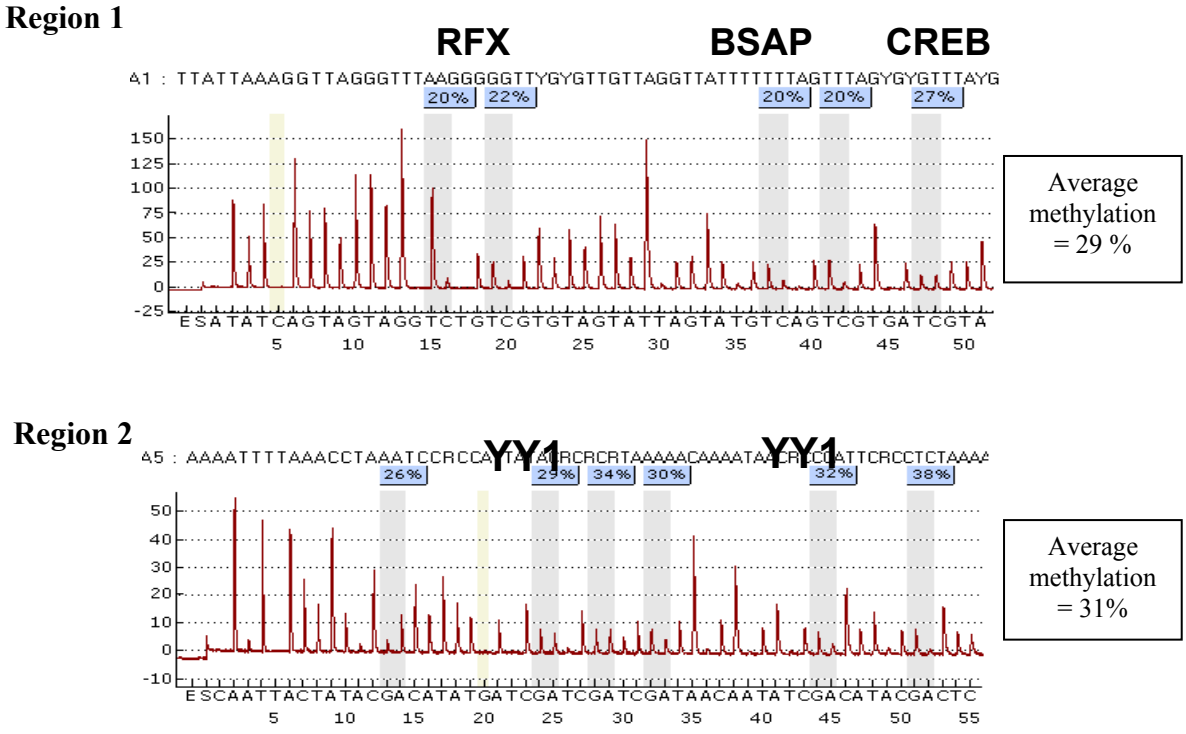



**Figure 5.12: Pyrosequencing results for the Wp promoter in the 2W EBV LCL.** The percentage methylation at each CpG is reported in the blue boxes above the program. The average methylation for the region is also reported.

In this section, I have shown that the Wp promoter is methylated and bound by DNMT3A in a region approximately 350bp up-stream of its transcription start site. The binding of DNMT to the Wp promoter in other EBV positive cell lines was broadly consistent with the previously reported methylation status of this promoter in these lines. Contrary to the impression left by an earlier report, I found that the Wp promoter is both bound by DNMT and methylated in the 2W EBV LCL.

Having shown in the preceding chapter that DNMT1 and DNMT3B are down-regulated in GC B cell derived LCLs. I have attempted to examine the impact on viral gene expression of re-expressing DNMT1 and DNMT3B in these cell lines.

## 5.9 Transient transfection of GC B derived LCLs with DNMT1 and DNMT3B using electroporation is followed by the up-regulation of these enzymes at the RNA level but not at the protein level

In the previous chapter, I showed that DNMT1 and DNMT3B are down-regulated in GC B cells following infection with EBV. Therefore, I next wanted to know what would be the effect on viral gene expression were I to re-express these proteins in GC B cell derived LCLs. Towards this end, I used pCDNA3-DNMT1 and pCDNA-DNMT3B plasmids provided by Dr. Keith Robertson (Department of Biochemistry and Molecular Biology, Shands Cancer Center, University of Florida). GC B cell derived LCLs (1 x$10^7$ cells) were transfected using electroporation with 10 μg, or 20 μg of pCDNA3-DNMT1, pCDNA-DNMT3B or vector control plasmid according to conditions specified in section 2.12. RNA and protein were extracted at 24 and 48 hours post transfection.

Q-RT PCR showed up-regulation of DNMT1 and DNMT3B 24 and 48 hours post transfection (Figure 5.13). However western blotting showed no up-regulation of DNMT1 or DNMT3B protein (Figure 5.14). The blots were re-probed for the plasmids HA tag but this too could not be detected. When cytospins were made from transfected cells as described in

section 2.14, immunohistochemistry revealed that only 1% of the cells had been successfully transfected compared to the vector control (Figure 5.15). When performing these experiments I recorded that there was approximately 50% cell death following transfection. Therefore I came to the conclusion that these cells could not tolerate the over-expression of DNMT1 or DNMT3B. Transfection experiments were repeated using Lipofectamine but with essentially identical results (data not shown).

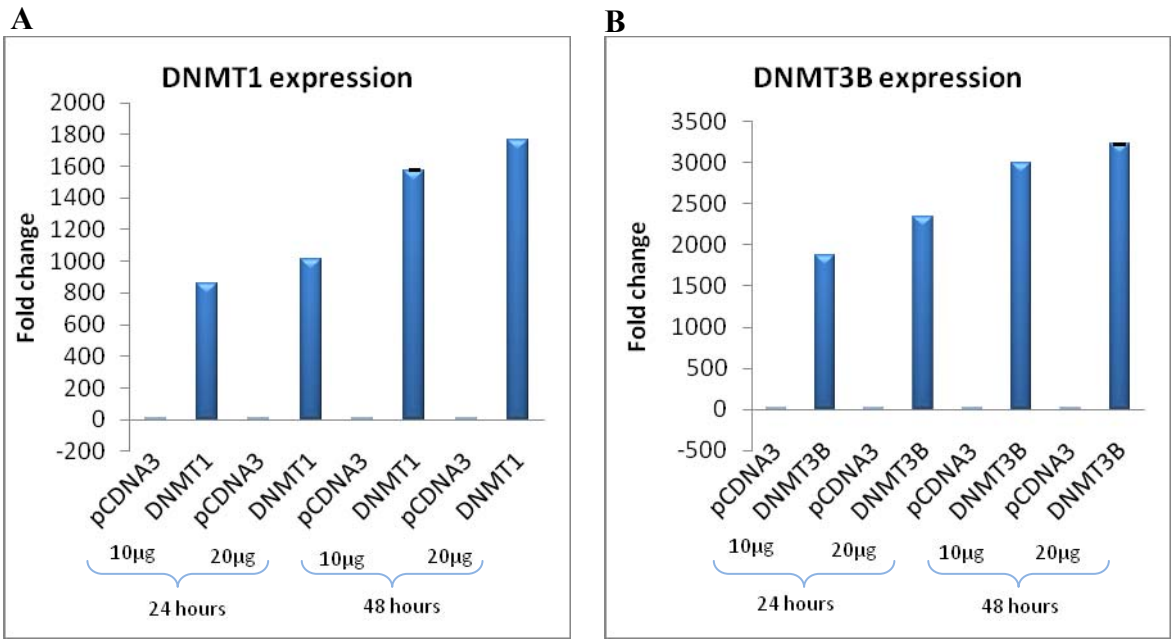



**Figure 5.13: DNMT1 and DNMT3B mRNA expression following transfection of GC B cell derived LCLs.** LCLs were transfected with 10µg or 20 µg of DNMT1 or DNMT3B and the expression was analysed 24 and 48 hours post transfection. All Q RT-PCR samples were analysed in triplicate and are presented as $2^{-\Delta\Delta CT}$ values compared with the vector control.

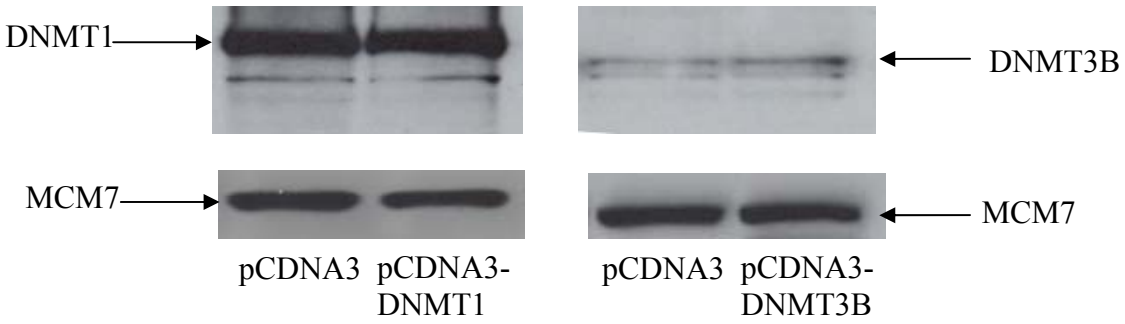



**Figure 5.14: Western blotting of DNMT1 and DNMT3B following electroporation.** DNMT1 and DNTM3B expression was detected using 20 µg of protein. MCM7 was used as a loading control.

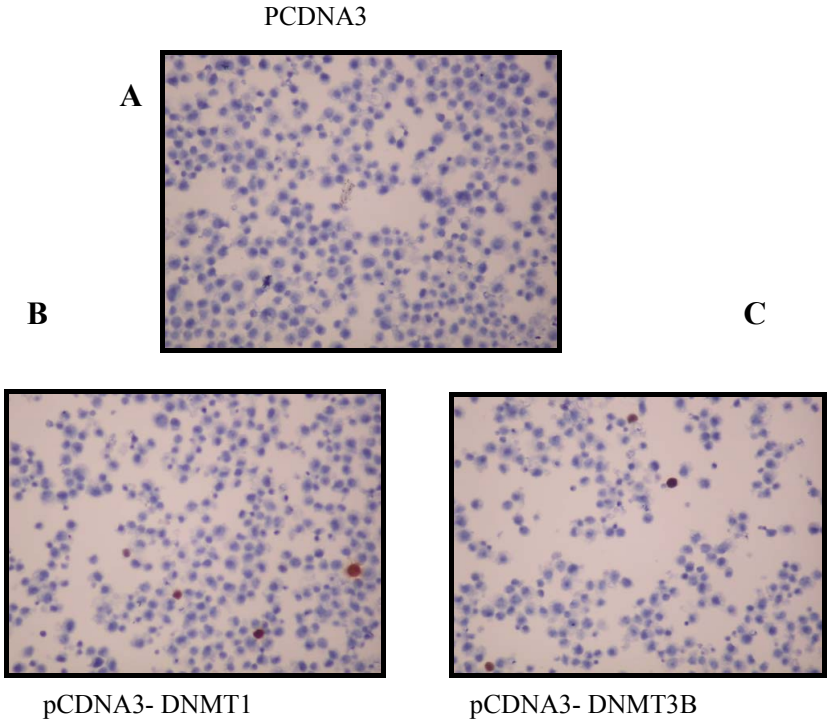



**Figure 5.15: DNMT1 and DNMT3B expression in GC B cell derived LCL.**

Immunohistochemistry for the HA tag expressed by LCLs successfully electroporated with pCDNA3- DNMT1 or pCDNA3- DNMT3B. **A.** Cells transfected with the pCDNA3 vector control. **B.** Cells transfected with pCDNA3-DNMT1. **C.** Cells transfected with pCDNA3-DNMT3B.

## 5.10 Transient transfection of the marmoset cell line B95.8 with DNMT1 and DNMT3B using electroporation is followed by the up-regulation of these enzymes at the RNA and protein level

I next explored the possibility that the pCDNA3-DNMT1 and pCDNA3-DNMT3B plasmids would be less "toxic" in another LCL. Towards this end, these plasmids were transfected using electroporation into the B95.8 cell line which I selected because I had previously shown that the pattern of DNMT expression in this cell line compared to GC B cells is similar to that observed in EBV infected GC B cells (Figure 5.16).

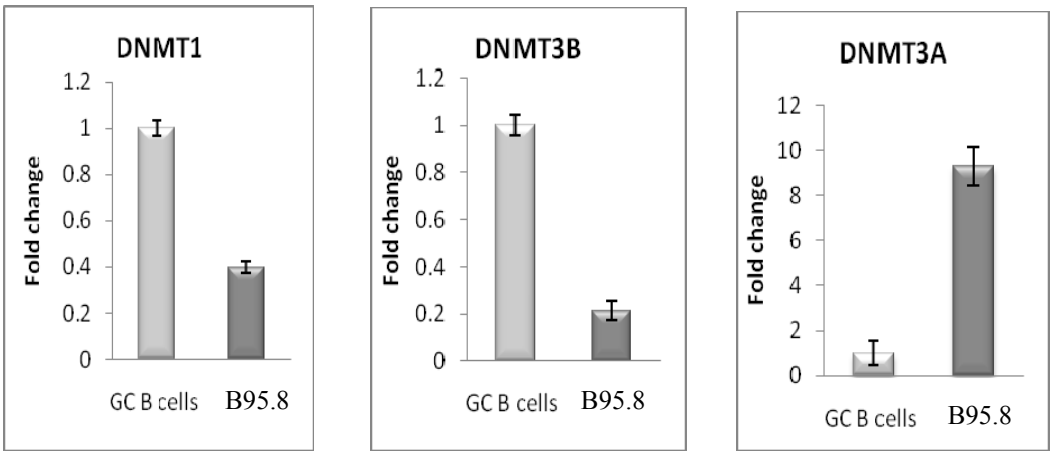



**Figure 5.16: DNMT1, DNMT3A and DNMT3B mRNA expression in B95.8 cells.** DNMT expression in B95.8 LCL was compared to GC B cells. Assays were carried out in triplicate and results are presented as $2^{-\Delta\Delta CT}$ values. Standard error bars show the variation in relative DNMT expression across three independent experiments.

Q-RT PCR showed up-regulation of DNMT1 and DNMT3B, 48 hours post transfection (Figure 5.17). On this occasion, western blotting showed an up-regulation of DNMT1 and DNMT3B at the protein level (Figure 5.18). When cytospins were made from transfected cells as described in section 2.14, immunohistochemistry confirmed the expression of HA-tagged DNMT1 and DNTM3B and revealed that approximately 10% of cells had been transfected (Figure 5.19).

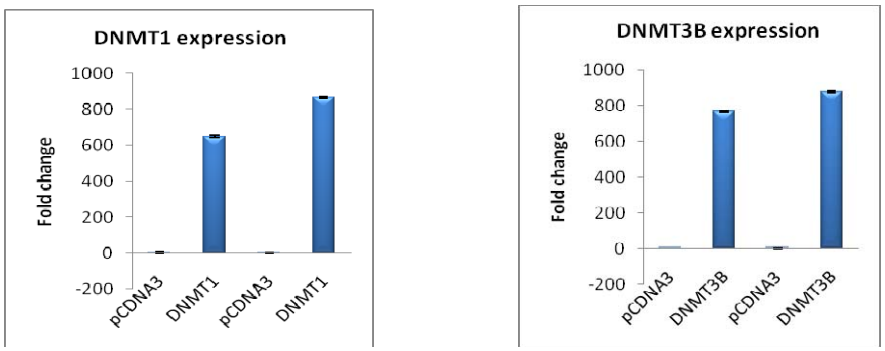



**Figure 5.17: DNMT1 and DNMT3B mRNA expression following transfection of B95.8 cells.** B95.8 cells were transfected with 10µg or 20 µg of DNMT1 or DNMT3B and the expression was analysed 48 hours post transfection. All Q RT-PCR samples were analysed in triplicate and are presented as $2^{-\Delta\Delta CT}$ values compared with the vector control.

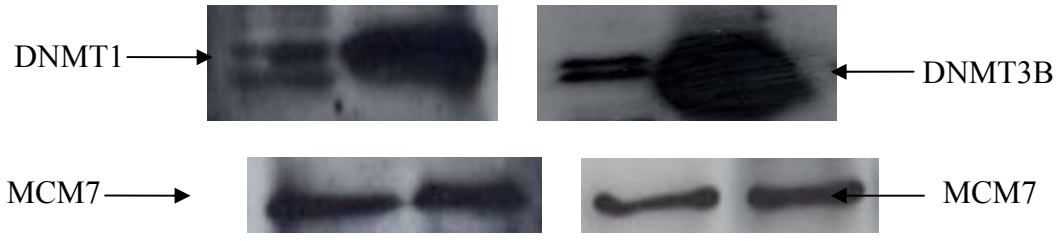



**Figure 5.18: Western blotting of DNMT1 and DNMT3B following electroporation into B95.8 cells.** DNMT1 and DNTM3B expression was detected using 20 µg of protein. MCM7 was used as a loading control.

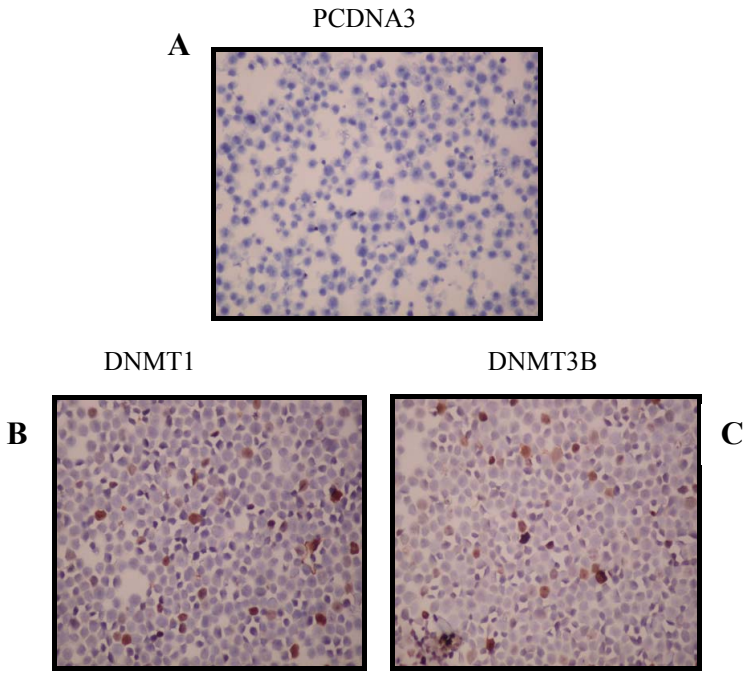
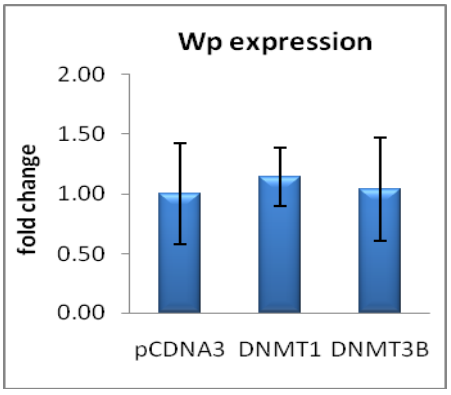


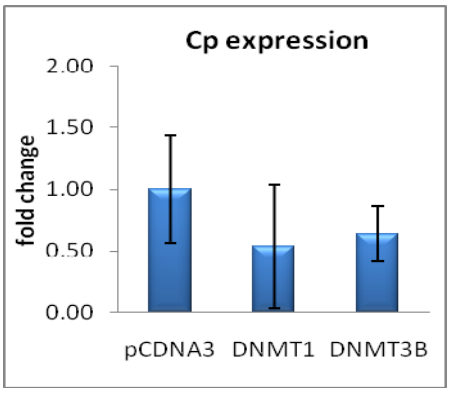

**Figure 5.19: DNMT1 and DNMT3B expression in B95.8 LCL.**

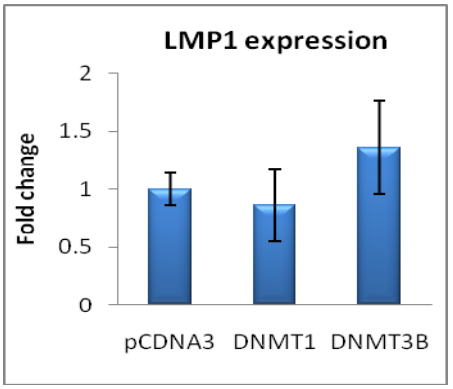
Immunohistochemistry for the HA tag expressed by LCLs successfully electroporated with pCDNA3- DNMT1 or pCDNA3- DNMT3B. **A.** Cells transfected with the pCDNA3 vector control (no HA tag expression). **B.** Cells transfected with pCDNA3-DNMT1. **C.** Cells transfected with pCDNA3-DNMT3B.

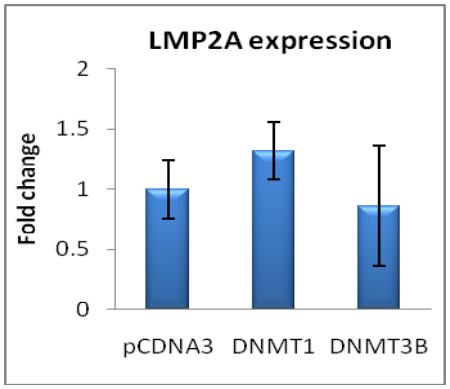
## 5.11 DNMT1 increases the expression of BZLF1 in B95.8 cells

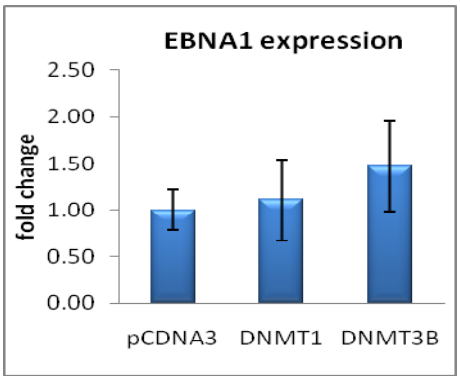
Having successfully transfected the B95.8 cells with DNMT1 and DNMT3B, I next explored whether over-expression of these enzymes affected viral gene expression. Q-RT PCR performed on each of the samples transfected with DNMT1 revealed up-regulation of BZLF1 compared to the vector control. There was no change in expression of BZLF1 following transfection with DNMT3B (Figure 5.20). The expression of LMP1, LMP2A, EBNA1, EBNA2, Wp, and Cp was unchanged following transfection with either DNMT1 or DNMT3B. I confirmed the increased expression of BZLF1 following transfection with DNMT1 using immunohistochemistry (Figure 5.21).

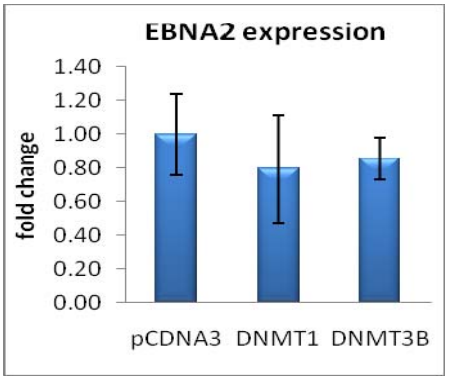

Wp expression
fold change
2.00
1.50
1.00
0.50
0.00
pCDNA3
DNMT1
DNMT3B


Cp expression
fold change
2.00
1.50
1.00
0.50
0.00
pCDNA3
DNMT1
DNMT3B


LMP1 expression
Fold change
2
1.5
1
0.5
0
pCDNA3
DNMT1
DNMT3B


LMP2A expression
Fold change
2
1.5
1
0.5
0
pCDNA3
DNMT1
DNMT3B


EBNA1 expression
fold change
2.50
2.00
1.50
1.00
0.50
0.00
pCDNA3
DNMT1
DNMT3B


EBNA2 expression
fold change
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
pCDNA3
DNMT1
DNMT3B

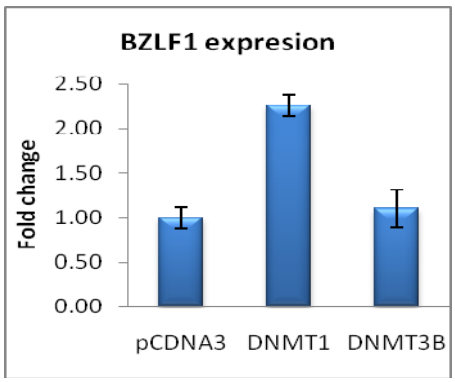

BZLF1 expresion
Fold change
2.50
2.00
1.50
1.00
0.50
0.00
pCDNA3
DNMT1
DNMT3B

**Figure 5.20: Expression of viral genes following DNMT1 and DNMT3B over-expression in B95.8 cells.** All Q RT-PCR samples were analysed in triplicate and are presented as $2^{-\Delta\Delta CT}$ values compared with the vector control.

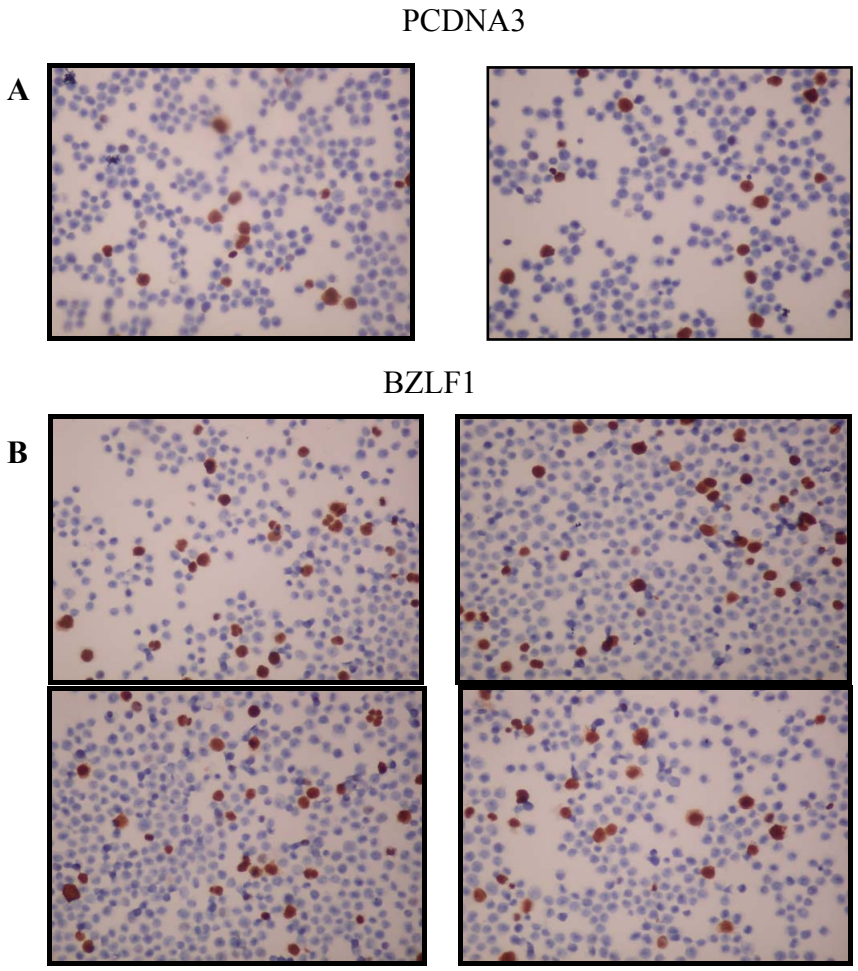



**Figure 5.21: BZLF1 expression in B95.8 cells transfected with DNMT1.** Cells expressing BZLF1 are stained brown, blue counterstained cells are negative for BZLF1 expression. **A.** Cells transfected with the pCDNA3 vector control expressing BZLF1. **B.** Cells transfected with pCDNA3-DNMT1 expressing BZLF1.

## 5.12 DNMT1 does not bind to the lytic genes, BZLF1, BRRF1 or BRLF1 in B95.8 cells

Having shown that BZLF1 was up-regulated following transfection of DNMT1, I next explored using X-ChIP whether this was associated with binding of DNMT1 to BZLF1. The X-ChIP analysis was extended to include the EBV lytic genes BRRF1 and BRLF1 which control the expression of BZLF1 and which themselves have been shown to be regulated by methylation (Dickerson *et al.,* 2009, Bhende *et al.,* 2005).

X-ChIP primers were designed to cover the BZLF1, BRRF1 and BRLF1 promoters and are shown in section 2.13.6 (page 85). Cells transfected with either DNMT1 or the vector control were harvested 48 hours post transfection and their chromatin prepared for X-ChIP as described in section 2.13. However I found no evidence of DNMT1 binding to the lytic genes BZLF1 (Figure 5.22), BRRF1 or BRLF1 (Figure 5.23).

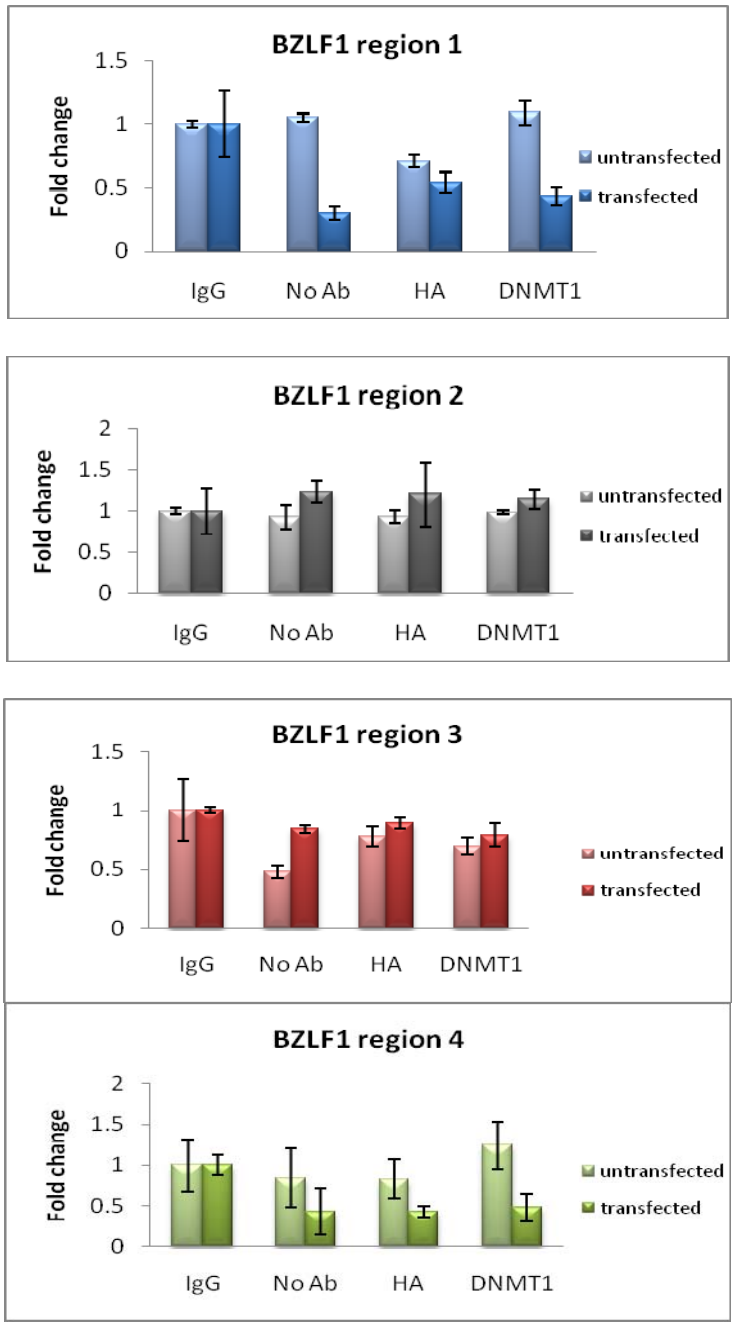



**Figure 5.22: SYBR green Q-PCR results of DNMT1 binding to BZLF1 by X-ChIP analysis.** DNMT1 binding to four regions of BZLF1 was assessed using X-ChIP. Transfected cells were compared to pCDNA3 vector control transfected cells. IgG and no antibody IPs were used as controls. Each assay was carried out in triplicate and standard error bars show the variation across the assays.

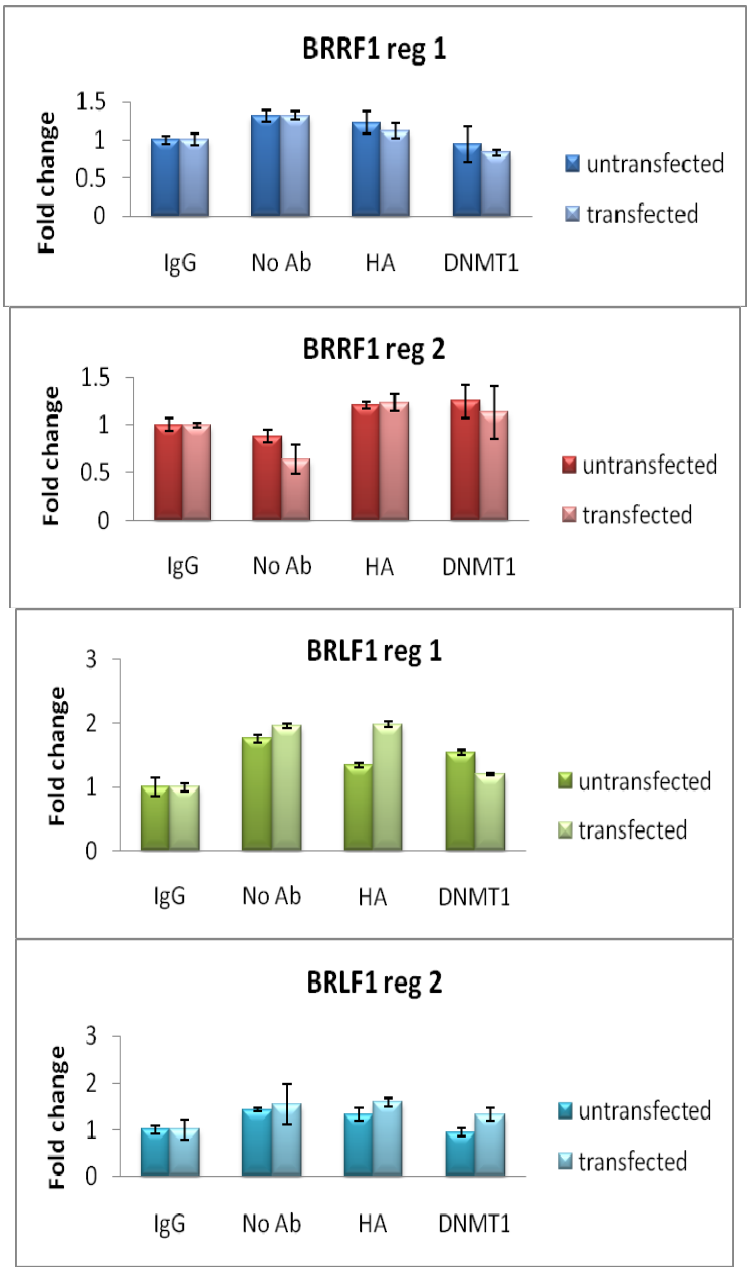
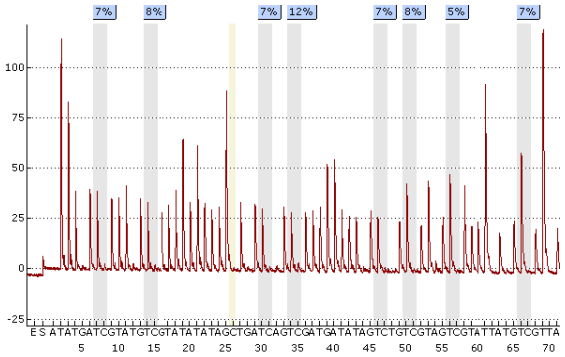



**Figure 5.23: SYBR green Q-PCR results of DNMT1 binding to BRRF1 or BRLF1 by X-ChIP analysis.** DNMT1 binding of two regions of BRRF1 and two regions of BRLF1 was assessed using X-ChIP. Transfected cells were compared to pCDNA3 vector control transfected cells. IgG and no antibody IPs were used as controls. Each assay was carried out in triplicate and standard error bars show the variation across the assays.

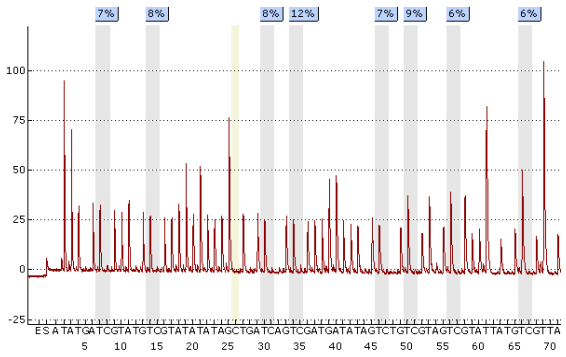
## 5.13 DNMT1 transfection of B95.8 cells is not followed by a change in the methylation of BZLF1, BRRF1 or BRLF1

Given that both the expression of BZLF1 and of those other lytic genes (BRRF1 and BRLF1) which control its expression are regulated by methylation, I next investigated whether DNMT1 transfection was followed by a change in methylation at the promoters of these genes. However, pyrosequencing revealed no change in methylation at the promoters of these genes following DNMT1 transfection (Figure 5.24).

**BZLF1**

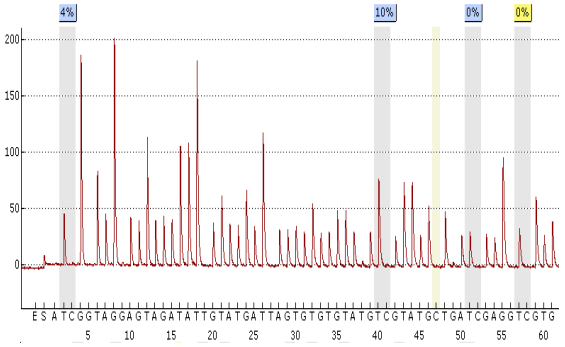



pCDNA3-vector only

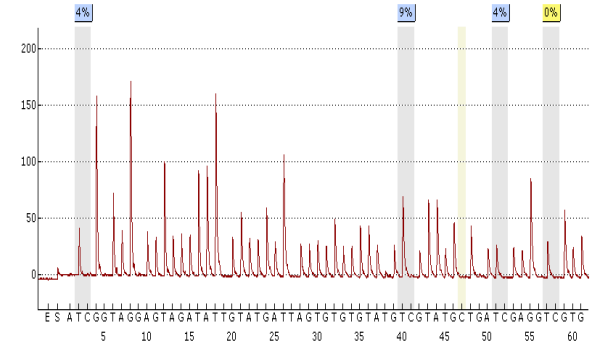



pCDNA3-DNMT1

**BRRF1**

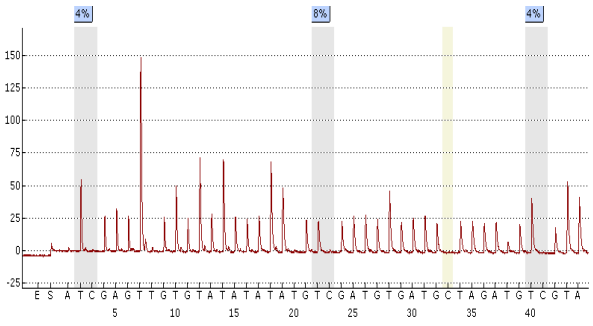



pCDNA3-vector only

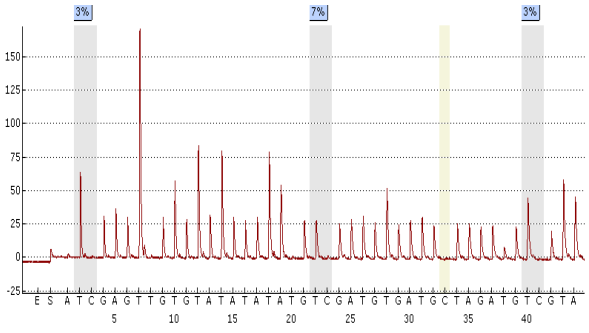



pCDNA3-DNMT1

**BRLF1**




pCDNA3-vector only




pCDNA3-DNMT1

**Figure 5.24: Pyrosequencing results for a region of the BZLF1, BRRF1 and BRLF1 promoters in B95.8 cells transfected with DNMT1.** The percentage methylation at each CpG is reported in the blue boxes above the program.

I have described in this section how I over-expressed DNMT1 and DNMT3B in the B95.8 cell line. Although I showed an up-regulation of BZLF1 following transfection with DNMT. I was not able to show DNMT1 binding to the BZLF1, BRLF1 or BRRF1 promoters. I was also unable to show a change in methylation at these promoters following DNMT1 transfection. Therefore I have no compelling evidence to suggest that DNMT1 directly regulates the expression of BZLF1.

Chapter 6

# Discussion

One of the difficulties faced when investigating the early events contributing to the pathogenesis of Hodgkin's lymphoma is the identification of an appropriate cell line model. Cell lines derived from lymphomas are already fully transformed, rarely resemble their progenitor cell and are likely to have culture-associated genetic and epigenetic changes. To investigate those early EBV-associated changes contributing to the pathogenesis of HL, I chose as my cell line model, an LCL derived from GC B cells, the presumed progenitor cells of HL. To determine the relevance of this model to the pathogenesis of HL, I compared the transcriptional profile of EBV-infected GC B cells with that of HL cell lines, micro-dissected HRS cells and GC B cells transfected with the EBV latent genes, LMP1 and LMP2A. Although I was obliged to compare the transcriptional profile of my GC B cell derived LCLs with that of GC B cells obtained from donors different from those initially infected with EBV, the results of this validation exercise were encouraging. Substantial and significant over-laps were found between the transcriptional profile of EBV infected GC B cells and those of micro-dissected HRS cells and of GC B cells transfected with LMP1, the major EBV oncogene. Perhaps the most compelling result was the finding that 80% of the genes down-regulated in HL cell lines and 76% of those up-regulated were concordantly regulated in EBV infected GC B cells. The strength and consistency of these associations suggest that EBV infected GC B cells are a useful model for investigating early virus associated epigenetic and transcriptional changes contributing to the pathogenesis of HL. As to whether I would see similar transcriptional changes if I compared my EBV infected GC B cell arrays to BL cell line arrays remains moot, as there are no BL arrays available for me to compare my arrays to.

In common with other oncogenic viruses, EBV de-regulates the expression of the DNA methyltransferases (Flanagan 2007). However, their pattern of expression in EBV infected GC B cells, is markedly different from that associated with other oncogenic viruses. Whereas these oncogenic viruses consistently up-regulate DNMT1, DNMT3A and DNMT3B, I found that EBV down-regulates DNMT1 and DNMT3B and up-regulates DNMT3A in GC B cells. I have also shown that LMP1, the major EBV oncogene, is responsible for the down-regulation of DNMT1. However, I could find no evidence to suggest that either DNMT3A or DNMT3B were down-regulated in GC B cells by the EBV latent genes, LMP1, EBNA1 or LMP2A. It is worth pointing out that EBV-induced changes in the expression of these epigenetic regulators are likely to be cell type-specific. DNMT1, DNMT3A and DNMT3B are up-regulated in nasopharyngeal cancer cell lines by the major EBV oncogene, LMP1, and DNMT1 is reported to be up-regulated in a gastric cancer cell line by another EBV latent gene, LMP2A (Tsai *et al.,* 2002, Seo *et al.,* 2008, Hino *et al.,* 2009). Furthermore, although DNMT1, DNMT3A and DNMT3B are consistently over-expressed in cancer at a wide range of sites, I found that DNMT1 and DNMT3B were down-regulated, and DNMT3A up-regulated in a panel of HL cell lines. This pattern of expression is consistent with the results of gene expression profiling performed on micro-dissected HRS cells (Brune *et al.,* 2008).

Although each of the HL cell lines showed an increase in DNMT3A and a decrease in DNMT3B and DNMT1 at the protein level, the transcriptional results did not always correlate. For example, DNMT3A was not up-regulated in L428 at the transcriptional level, this could point to a possible role for post-transcriptional modifications such as miRNA regulation of DNMTA in this cell line, or it could also be due to a mutation within the region being analysed by Q-RT-PCR, which could be checked by sequencing the region covered by the Q- RT PCR primers.

My observations offer some clues as to how EBV-induced de-regulation of the DNA methyltransferases might contribute to viral persistence and to the pathogenesis of HL. For example, DNMT1 has been shown to have an essential role in maintaining the progenitor state of constantly replenishing somatic tissue (Sen *et al.,* 2010). Depletion of DNMT1 is followed by the exit of cells from the progenitor compartment and their premature differentiation. This transition is associated with the down-regulation of UHRF1, a component of the cell's DNA methylation machinery, and with the up-regulation of GADD45, a putative DNA demethylase. Similar transcriptional changes are seen during normal B cell differentiation when DNMT1 and UHRF1 are down-regulated and GADD45A up-regulated in both plasma cells and memory cells compared with centrocytes (Brune *et al.,* 2008, supplemental information). I have confirmed that DNMT1 expression falls in B cell subsets with increasing stage of differentiation and have shown using gene expression profiling that the down-regulation of DNMT1 observed in EBV infected GC B cells is associated with both the down-regulation of UHRF1 (fold change, 2.8) and with the up-regulation of GADD45A (fold change, 2.4) Therefore, LMP1-induced down-regulation of DNMT1 in GC B cells may explain in part how this viral oncogene drives cells towards a post GC stage while at the same time hijacking the B cell transcriptional programme and subverting normal B cell differentiation. These are important considerations given that DNA demethylating agents which deplete DNMT1 are already in Phase II trials for the treatment of HL. (http://forums.lymphoma.com/archive/index.php/t-24166.html).

My observations also reveal how EBV-induced changes in the expression of the DNMTs might contribute to the establishment of persistent viral infection. Although Wp is the first viral promoter to be activated during the *in vitro* transformation of primary resting B cells, levels of Wp-initiated transcripts decline and Cp becomes the dominant EBNA promoter in most established LCL. This flexibility of latent promoter usage has been considered critical to the virus' strategy for persistence in vivo (Hutchings *et al.,* 2006). While it has been known for some time that Wp is silenced by DNA methylation, I was able to reveal a potential role for EBV induced up-regulation of DNMT3A in maintaining viral persistence by demonstrating that this enzyme binds to and is associated with methylation of the Wp promoter. However, I was less successful in showing how EBV-induced down-regulation of DNMT1 and DNMT3B might contribute to the viral life cycle. Although I found that over-expression of DNMT1 in B95.8 cells was followed by the up-regulation of the early lytic gene BZLF1, I was unable to show DNMT1 binding to the BZLF1promoter. Nor was I was able to show that over-expression of DNMT1 was followed by a change in methylation status at either the BZLF1 promoter or those of the other lytic cycle genes, BRRF1 and BZLF1 which regulate its expression. Therefore, I have to conclude that the observed up-regulation of BZLF1 which I observed is not a direct consequence of the over-expression of DNMT1.

Turning now to the methylation arrays which I performed on EBV infected GC B cells. As this was the first time methylation arrays had been performed in the School of Cancer Sciences, and no bespoke method was available for their analysis, I was fortunate to have the assistance of Dr Wenbin Wei, Head of Bioinformatics in this School. After much consideration and some preliminary validation studies, we decided to analyse my arrays using the Tilemap software programme and the moving average method. This software has since been used by others when analysing tiling arrays (Akkers *et al.,* 2010, Zinzen *et al.,* 2009,

Reed *et al.,* 2008). Although I found that pyrosequencing confirmed almost all the changes predicted on my methylation array, the magnitude of these changes was often surprisingly small. This is perhaps inevitable given that methylation arrays are performed on methyl-enriched samples whereas pyrosequencing is performed on un-enriched DNA. Whereas the former method selects only those cells which are methylated, the latter must detect methylation changes which are dispersed through the total cell population. An alternative way to confirm changes predicted on the methylation array would have been to methyl-enrich the sample and then to look for methylated forms using Q-PCR.

EBV-induced changes in the expression of the DNMT in GC B cells were associated with methylation changes in a substantial number of cellular genes. However, these changes were not randomly distributed across the genome but appeared to cluster at certain chromosomal locations. Concordant methylation of adjacent CpG island gene promoters has also been reported for a number of gene clusters in cancer, and recent genome-wide analyses have identified other large chromosomal regions containing several CpG islands which are often methylated and transcriptionally repressed in cancer (Coolen *et al.,* 2010), Although these observations suggest that coordinated epigenetic control over large regions is common in cancer, this phenomenon has not previously been reported following the infection of primary cells with an oncogenic virus (Davidsson *et al.,* 2009, Taylor *et al.,* 2007).

As to why some genes become more or less methylated following EBV infection of GC B cells, and others not, also appear to depend on their CpG content. Whereas promoters with a high CpG content were significantly more likely to become methylated following EBV infection, those with a low CpG content were more likely to become less methylated This relationship between CpG content and methylation change has been described before and appears to be context dependent. For example, in melanoma, prostate cancer, and in non-Hodgkin lymphomas, the trend is the same as that observed following EBV infection: genes with a high CpG content are more likely to have increased methylation whereas those with a low CpG content are more likely to be hypomethylated (Koga *et al.,* 2009, Martin-Subero *et al.,* 2009, Gal-Yam *et al.,* 2008). In contrast, Weber *et al.,* 2007 found that during normal development, promoters with a low CpG content were more likely to be targeted by methylation whereas those with a high CpG content remained un-methylated.

Subsequent to undertaking the work presented in this thesis, I have performed methylation arrays on foreskin keratinocytes transfected with episomal HPV16. The analysis of these arrays showed that the relationship between CpG content and the direction of HPV-induced methylation changes is almost identical to the pattern observed following EBV infection of GC B cells. Thus it would appear that in cancer and in primary cells exposed to oncogenic viruses*, de novo* methylation changes are determined in part by cis-acting susceptibility factors such as promoter CpG content.

One of the objectives of my thesis was to determine how often EBV-associated methylation events predicted virus-associated transcriptional changes. My results were perhaps surprising given the widely held belief that an increase in methylation is inevitably associated with transcriptional down-regulation, and a decrease in methylation with transcriptional up-regulation. In fact, I found that genes with increased methylation following EBV infection of GC B cells were significantly more likely to be either transcriptionally increased, whereas those with decreased methylation were significantly less likely to be changed in either direction. Others who have performed integrated methylation and transcriptional profiling have concluded that the relationship between methylation and transcriptional ***change*** is more complex than was previously thought (Boellmann *et al.,* 2010, Shann *et al.,* 2008,). In my introduction, I have provided specific examples of where methylated sequences can facilitate and inhibit the binding of transcription factors. I now discuss more generally the interpretation of integrated epigenetic-transcriptional profiling studies.

One of the problems when attempting to correlate methylation and transcriptional changes is that these changes cannot be linked at the level of the individual cell. There is no way of knowing which cells have been affected by methylation change and which have undergone a transcriptional change. Although pyrosequencing can provide an estimate of the proportion of cells which contain methylated forms of a gene, transcriptional read-outs can only provide a summary measure across all cells in the sample, one which necessarily reflects the net effect of competing regulatory pressures within individual cells and across different cells. The kinetics of methylation induced transcriptional change are likely to be complex, and to be both gene and context dependent. Consider for example, PRDM1, the master regulator of plasma cell differentiation. A fellow PhD student (Katerina Vrzalicova) is investigating how EBV-induced changes in the expression of this gene contributes to abnormal B cell

differentiation. This gene has an alpha isoform and a beta isoform which is normally silenced in GC B cells. My methylation array predicted that the beta isoform became less methylated following EBV infection of GC B cells, a prediction I confirmed using pyrosequencing. Whereas this isoform was found to be methylated in 28% of uninfected GC B cell, the proportion of infected cells fell to 10 % six weeks following infection with EBV, at which point Katerina was able to show using Q-RT-PCR that its expression had been induced. For this particular isoform in this cell background, an 18% reduction in the number of cells containing methylated forms was sufficient to induce gene expression. For another gene, the fall in the number of cells containing methylated form necessary to trigger transcription would almost certainly be different. This example illustrates another difficulty which may arise when attempting to correlate methylation and transcriptional changes using observations based on only two time-points. I have shown that EBV associated methylation changes in GC B cells are progressive, probably as a consequence of a monoclonal population growing out and becoming polyclonal. It follows from this that observations made earlier than six weeks might have revealed methylation changes in a particular gene but failed to reveal methylation-associated transcriptional changes because sufficient cells had yet to acquire methylated forms.

When attempting to interpret the results of integrated methylation-transcriptional profiling, there are also technical constraints relating to the coverage and resolution provided by the array platforms. These may frustrate attempts to reveal the true relationship between methylation and transcriptional changes. For example, when I started my thesis, the emphasis was on detecting methylation changes in or immediately adjacent to the gene promoter. Although the Affymetrix platform used in my experiments provides extensive coverage of gene promoters, this approximates to only 7% of the genome. This is an important

consideration given that it has recently been shown that intragenic methylation may be an important determinant of transcriptional change (Maunakea *et al.,* 2010). Furthermore the U133 Affymetrix array used in my transcriptional analysis does not have probes for all known isoforms of each gene. For example, whereas it was possible to distinguish methylation changes in the different isoforms of the PRDM1 gene on the Affymetrix promoter array, only probes for the alpha isoform and not the beta isoform are included on the transcriptional array. Although I found that only one TSG (TP73) silenced by DNA methylation in HL was also methylated in GC B cells by EBV, my literature search identified only twenty-two epigenetically silenced TSG in HL, and no instance of a gene being hypomethylated in this lymphoma. In these circumstances, evidence of absence may simply reflect absence of evidence. For these reasons, a final judgment on the relevance of the early methylation changes revealed on my GC B cell derived LCL to the pathogenesis of HL must be postponed until we have the results of a comparable methylation array performed on micro-dissected HRS cells. Performing methylation arrays on HRS cells would be technically challenging given the number of cells that would need to be micro-dissected in order to perform a MeDIP. However, technical advances now make it possible to perform these assays on much smaller cell numbers than before (Active Motif, MethylCollector™ Ultra).

Finally, one of the unexpected findings on my transcriptional array was the up-regulation of genes that modulate another epigenetic mark, arginine methylation, In a series of preliminary experiments, I have confirmed that EBV infection of GC B cells, the presumptive progenitors of the HRS cells found in HL, is followed by the up-regulation of the protein arginine methyltransferases CARM1, PRMT1 and PRMT5, and the down-regulation of the arginine deiminase, PADI4. This pattern of expression is recapitulated in HL cell lines and in primary HL, and is entirely consistent with the results of gene expression profiling performed on micro-dissected HRS cells (Brune *et al.,* 2008, supplemental information). My finding that the expression of CARM1 was increased at the protein level in the HL cell line L450 but decreased at the RNA level points to a possible role for post-transcriptional modification in the regulation of the arginine methyltransferases. Increased protein expression of PRMT5 has also been reported in transformed lymphoid cells despite levels of mRNA substantially less than those observed in normal cells (Pal *et al.,* 2007). This discrepancy has been causally attributed to the down-regulation of two miRNA, one of which, mir-96, has also been shown to be selectively down-regulated in EBV positive HL (Navarro *et al.,* 2008). EBV-induced changes in the expression of the PRMTs may also contribute to the viral life cycle. For example both PRMT1 and PRMT5 bind to the Epstein-Barr virus protein EBNA1 which is important for the replication and mitotic segregation of viral genomes, and which is re-localised following inhibition of PRMT1 (Shire *et al.,* 2006). Arginine dimethylation of another EBV oncoprotein, EBNA2, is necessary for its efficient association with DNA bound transcription factors and with other viral promoters (Gross *et al.,* 2010).

My observations offer some insights as to how EBV-induced changes in the expression of the PRMT might contribute to disturbances of B cell differentiation. Signals processed through the B cell antigen receptor control both proliferation and differentiation. PRMT,1 which I

have shown to up-regulated by LMP1 in GC B cells, methylates a conserved arginine in the CD79A subunit of the B cell receptor (BCR). This modification negatively regulates the calcium and PI-3 kinase pathways in the BCR while promoting signals leading to B cell differentiation (Infantino *et al.,* 2010). LMP1 has also been shown to increase the storage of Ca++ in the endoplasmic reticulum and to down-regulate the PI-3 kinase pathway in B cells (Dellis *et al*., 2009). Given that our laboratory has shown that LMP1 can drive cells towards a post GC stage of differentiation, it will be important to investigate further the contribution of virus-induced de-regulation of PRMT1 to abnormal B cell differentiation (Vockerodt *et al.,* 2008). Additional evidence linking EBV induced de-regulation of the PRMTs to the pathogenesis of HL is possibly provided by the observation that knockdown of PRMT5 reduces proliferation in transformed B cells and restores expression of RBL2. We have found using gene expression profiling that LMP1 down regulates this pocket protein in GC B cells, and we and others have shown that it is also down-regulated in both HL cell lines and in micro-dissected HRS cells (Wang *et al.,* 2008, Vockerodt *et al.,* 2008, supplemental information, Brune *et al.,* 2008, supplemental information).

More than thirty years ago inhibition of PRMT1 was shown to inhibit Rous sarcoma virus induced chick embryo fibroblast transformation (Enouf *et al.,* 1979). More recently an inhibitor of arginine methyltranferases was shown to reduce Tax transactivation in HTLV-1 transformed cells while at the same reducing NF Kappa B activity and inducing apoptosis (Dasgupta *et al.,* 2008). The continuing intensive effort to identify specific arginine methylation provides compelling reasons for endeavouring to dissect the contribution to transformation of virus induced changes in those proteins which regulate arginine methylation.

# Bibliography

Akkers RC, van Heeringen SJ, Manak JR, Green RD, Stunnenberg HG, Veenstra GJ (2010). ChIP-chip designs to interrogate the genome of Xenopus embryos for transcription factor binding and epigenetic regulation. *PLoS One* **5**: 8820-8831.

Alfieri C, Birkenbach M, Kieff E (1991). Early events in Epstein-Barr virus infection of human B lymphocytes. *Virology* **181:** 595-608.

Allan GJ, Inman GJ, Parker BD, Rowe DT, Farrell PJ (1992). Cell growth effects of Epstein-Barr virus leader protein. *J Gen Virol* **73:** 1547-1551.

Allan GJ, Rowe DT (1989). Size and stability of the Epstein-Barr virus major internal repeat (IR-1) in Burkitt's lymphoma and lymphoblastoid cell lines. *Virology* **173:** 489-98.

Allday MJ, Kundu D, Finerty S, Griffin BE (1990). CpG methylation of viral DNA in EBV-associated tumours. *Int J Cancer* **45:** 1125-1130.

Ambinder RF, Robertson KD, Tao Q (1999). DNA methylation and the Epstein-Barr virus. *Semin Cancer Biol* **9:** 369-375.

Ancelin K, Lange UC, Hajkova P, Schneider R, Bannister AJ, Kouzarides T, Surani MA (2006). Blimp1 associates with Prmt5 and directs histone arginine methylation in mouse germ cells. *Nat Cell Biol* **8:** 623-630.

Appanah R, Dickerson DR, Goyal P, Groudine M, Lorincz MC (2007). An unmethylated 3' promoter-proximal region is required for efficient transcription initiation. *PLoS Genet* **3:** 27-34.

Barreto G, Schäfer A, Marhold J, Stach D, Swaminathan SK, Handa V, Döderlein G, Maltry N, Wu W, Lyko F, Niehrs C (2007). Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. *Nature* **445:** 671-675.

Bechtel D, Kurth J, Unkel C, Kuppers R (2005). Transformation of BCR-deficient germinal-center B cells by EBV supports a major role of the virus in the pathogenesis of Hodgkin and posttransplantation lymphomas. *Blood* **106:** 4345-4350.

Bedford MT, Clarke SG (2009). Protein arginine methylation in mammals: who, what, and why. *Mol Cell* **33:** 1-13.

Bedford MT, Richard S (2005). Arginine methylation an emerging regulator of protein function. *Mol Cell* **18:** 263-272.

Ben-Sasson SA, Klein G (1981). Activation of the Epstein-Barr virus genome by 5-aza-cytidine in latently infected human lymphoid lines. *Int J Cancer* **28:**131-5.

Bestor TH (2005). Transposons reanimated in mice. *Cell* **122:** 322-325.

Bhende PM, Seaman WT, Delecluse HJ, Kenney SC (2004). The EBV lytic switch protein, Z, preferentially binds to and activates the methylated viral genome. *Nat Genet* **36:** 1099-1104.

Bhende PM, Seaman WT, Delecluse HJ, Kenney SC (2005). BZLF1 activation of the methylated form of the BRLF1 immediate-early promoter is regulated by BZLF1 residue 186. *J Virol* **79**: 7338-7348.

Biggin M, Bodescot M, Perricaudet M, Farrell P (1987). Epstein-Barr virus gene expression in P3HR1-superinfected Raji cells. *J Virol* **61:** 3120-3132.

Bird A (2001). Methylation talk between histones and DNA. *Science* **294:** 2113-2115.

Bird A (2002). DNA methylation patterns and epigenetic memory. *Genes Dev* **16:** 6-21.

Boellmann F, Zhang L, Clewell HJ, Schroth GP, Kenyon EM, Andersen ME, Thomas RS (2010). Genome wide analysis of DNA methylation and gene expression changes in the mouse lung following subchronic arsenate exposure. *Toxicol Sci* Epub ahead of print.

Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH (2001). Dnmt3L and the establishment of maternal genomic imprints. *Science* **294:** 2536-2539.

Brune V, Tiacci E, Pfeil I, Döring C, Eckerle S, van Noesel CJ, Klapper W, Falini B, von Heydebreck A, Metzler D, Bräuninger A, Hansmann ML, Küppers R (2008). Origin and pathogenesis of nodular lymphocyte-predominant Hodgkin lymphoma as revealed by global gene expression analysis. *J Exp Med* **205**: 2251-2268.

Burgers WA, Blanchon L, Pradhan S, de Launoit Y, Kouzarides T, Fuks F (2007). Viral oncoproteins target the DNA methyltransferases. *Oncogene* **26:** 1650-1655.

Butcher DT, Rodenhiser DI (2007). Epigenetic inactivation of BRCA1 is associated with aberrant expression of CTCF and DNA methyltransferase (DNMT3B) in some sporadic breast tumours. *Eur J Cancer* **43:** 210-219.

Caldwell RG, Wilson JB, Anderson SJ, Longnecker R (1998). Epstein-Barr virus LMP2A drives B cell development and survival in the absence of normal B cell receptor signals. *Immunity* **9**: 405-411.

Cai X, Schäfer A, Lu S, Bilello JP, Desrosiers RC, Edwards R, Raab-Traub N, Cullen BR (2006). Epstein-Barr virus microRNAs are evolutionarily conserved and differentially expressed. *PLoS Pathog* **23.**

Chan AT, Tao Q, Robertson KD, Flinn IW, Mann RB, Klencke B, Kwan WH, Leung TW, Johnson PJ, Ambinder RF (2004). Azacitidine induces demethylation of the Epstein-Barr virus genome in tumors. *J Clin Oncol* **22:** 1373-1381.

Chen J, Ueda K, Sakakibara S, Okuno T, Parravicini C, Corbellino M, Yamanishi K (2001). Activation of latent Kaposi's sarcoma-associated herpesvirus by demethylation of the promoter of the lytic transactivator. *Proc Natl Acad Sci U S A* **98:** 4119-4124.

Chen T, Hevi S, Gay F, Tsujimoto N, He T, Zhang B, Ueda Y, Li E (2007). Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells. *Nat Genet* **39:** 391-396.

Chen T, Ueda Y, Dodge JE, Wang Z, Li E (2003). Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol Cell Biol* **23:** 5594-5605.

Cheng D, Côté J, Shaaban S, Bedford MT (2007). The arginine methyltransferase CARM1 regulates the coupling of transcription and mRNA processing. *Mol Cell* **25:** 71-83.

Cohen JI, Lekstrom K (1999). Epstein-Barr virus BARF1 protein is dispensable for B-cell transformation and inhibits alpha interferon secretion from mononuclear cells. *J Virol* **73:** 7627-7632.

Coolen MW, Stirzaker C, Song JZ, Statham AL, Kassir Z, Moreno CS, Young AN, Varma V, Speed TP, Cowley M, Lacaze P, Kaplan W, Robinson MD, Clark SJ (2010). Consolidation of the cancer genome into domains of repressive chromatin by long-range epigenetic silencing (LRES) reduces transcriptional plasticity. *Nat Cell Biol* **12**: 235-246.

Dacwag CS, Ohkawa Y, Pal S, Sif S, Imbalzano AN (2007). The protein arginine methyltransferase Prmt5 is required for myogenesis because it facilitates ATP-dependent chromatin remodeling. *Mol Cell Biol* **27:** 384-394.

Dalla-Favera R, Lombardi L, Pelicci PG, Lanfrancone L, Cesarman E, Neri A (1987). Mechanism of activation and biological role of the c-myc oncogene in B-cell lymphomagenesis. *Ann N Y Acad Sci* **511:** 207-218.

Dasgupta A, Jung KJ, Jeong SJ, Brady JN (2008). Inhibition of methyltransferases results in induction of g2/m checkpoint and programmed cell death in human T-lymphotropic virus type 1-transformed cells. *J Virol* **82:** 49-59.

Daskalos A, Nikolaidis G, Xinarianos G, Savvari P, Cassidy A, Zakopoulou R, Kotsinas A, Gorgoulis V, Field JK, Liloglou T (2009). Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer. *Int J Cancer* **124:** 81-81.

Davidsson J, Lilljebjörn H, Andersson A, Veerla S, Heldrup J, Behrendtz M, Fioretos T, Johansson B (2009). The DNA methylome of pediatric acute lymphoblastic leukemia. *Hum Mol Genet* **18**: 4054-4065.

Dellis O, Arbabian A, Brouland JP, Kovàcs T, Rowe M, Chomienne C, Joab I, Papp B (2009). Modulation of B-cell endoplasmic reticulum calcium homeostasis by Epstein-Barr virus latent membrane protein-1. *Mol Cancer.* **8:** 59-68.

Deng T, Kuang Y, Wang L, Li J, Wang Z, Fei J (2009). An essential role for DNA methyltransferase 3a in melanoma tumorigenesis. *Biochem Biophys Res Commun* **387:** 611-616.

Di Bartolo DL, Cannon M, Liu YF, Renne R, Chadburn A, Boshoff C, Cesarman E (2008). KSHV LANA inhibits TGF-beta signaling through epigenetic silencing of the TGF-beta type II receptor. *Blood* **111:** 4731-4740.

Dickerson SJ, Xing Y, Robinson AR, Seaman WT, Gruffat H, Kenney SC (2009). Methylation-dependent binding of the epstein-barr virus BZLF1 protein to viral promoters. *PLoS Pathog* **5**: 1000356-1000364.

Diehl V, Kirchner HH, Burrichter H, Stein H, Fonatsch C, Gerdes J, Schaadt M, Heit W, Uchanska-Ziegler B, Ziegler A, Heintz F, Sueno K (1982). Characteristics of Hodgkin's disease-derived cell lines. *Cancer Treatment Reports* **66**: 615-632.

Doerr JR, Malone CS, Fike FM, Gordon MS, Soghomonian SV, Thomas RK, Tao Q, Murray PG, Diehl V, Teitell MA, Wall R (2005). Patterned CpG methylation of silenced B cell gene promoters in classical Hodgkin lymphoma-derived and primary effusion lymphoma cell lines. *J Mol Biol* **350:** 631-640.

Dong CW, Zhang YB, Lu AJ, Zhu R, Zhang FT, Zhang QY, Gui JF (2007). Molecular characterisation and inductive expression of a fish protein arginine methyltransferase 1 gene in response to virus infection. *Fish Shellfish Immunol* **22:** 380-393.

Dressler GR, Rock DL, Fraser NW (1987). Latent Herpes Simplex Virus Type 1 DNA Is Not Extensively Methylated *in vivo*. *J gen Virol* **68:** 1761-1765.

Drexler HG, Leber BF, Norton J, Yaxley J, Tatsumi E, Hoffbrand AV, Minowada J (1988). Genotypes and immunophenotypes of Hodgkin's disease-derived cell lines. *Leukemia* **2**: 371-376.

Eden A, Gaudet F, Waghmare A, Jaenisch R (2003). Chromosomal instability and tumors promoted by DNA hypomethylation. *Science* **300:** 455-463.

Egger G, Jeong S, Escobar SG, Cortez CC, Li TW, Saito Y, Yoo CB, Jones PA, Liang G (2006). Identification of DNMT1 (DNA methyltransferase 1) hypomorphs in somatic

knockouts suggests an essential role for DNMT1 in cell survival. *Proc Natl Acad Sci U S A* **103:** 14080-14085.

Ehlers A, Oker E, Bentink S, Lenze D, Stein H, Hummel M (2008). Histone acetylation and DNA demethylation of B cells result in a Hodgkin-like phenotype. *Leukemia* **22:** 835-841.

Enouf J, Lawrence F, Tempete C, Robert-Gero M, Lederer E (1979). Relationship between inhibition of protein methylase I and inhibition of Rous sarcoma virus-induced cell transformation. *Cancer Res* **39:** 4497-4502.

Epstein MA, Achong BG (1979). The Epstein-Barr virus. Springer-Verlag, Berlin, Heildelberg, New York.

Ernberg I, Falk K, Minarovits J, Busson P, Tursz T, Masucci MG, Klein G (1989). The role of methylation in the phenotype-dependent modulation of Epstein-Barr nuclear antigen 2 and latent membrane protein genes in cells latently infected with Epstein-Barr virus. *J Gen Virol* **70:** 2989-3002.

Esteller M (2005). Aberrant DNA methylation as a cancer-inducing mechanism. *Annu Rev Pharmacol* Toxicol **45:** 629-656.

Etoh T, Kanai Y, Ushijima S, Nakagawa T, Nakanishi Y, Sasako M, Kitano S, Hirohashi S (2004). Increased DNA methyltransferase 1 (DNMT1) protein expression correlates significantly with poorer tumor differentiation and frequent DNA hypermethylation of multiple CpG islands in gastric cancers. *Am J Pathol* **164:** 689-699.

Fabbrizio E, El Messaoudi S, Polanowska J, Paul C, Cook JR, Lee JH, Negre V, Rousset M, Pestka S, Le Cam A, Sardet C (2002). Negative regulation of transcription by the type II arginine methyltransferase PRMT5. *EMBO Rep* **3:** 641-645.

Falk KI, Szekely L, Aleman A, Ernberg I (1998). Specific methylation patterns in two control regions of Epstein-Barr virus latency: the LMP-1-coding upstream regulatory region and an origin of DNA replication (oriP). *J Virol* **72:** 2969-2974.

Falk KI, Zou JZ, Lucht E, Linde A, Ernberg I (1997). Direct identification by PCR of EBV types and variants in clinical samples. *J Med Virol* **51:** 355-363.

Falk MH, Tesch H, Stein H, Diehl V, Jones DB, Fonatsch C, Bornkamm GW (1987). Phenotype versus immunoglobulin and T-cell receptor genotype of Hodgkin-derived cell lines: activation of immature lymphoid cells in Hodgkin's disease. *Int J Cancer* **40**: 262-269.

Farina A, Feederle R, Raffa S, Gonnella R, Santarelli R, Frati L, Angeloni A, Torrisi MR, Faggioni A, Delecluse HJ (2005). BFRF1 of Epstein-Barr virus is essential for efficient primary viral envelopment and egress. *J Virol* **79:** 3703-3712.

Farrell, C.J., Lee, J.M., Shin, E.C., Cebrat, M., Cole, P.A. and Hayward, S.D. (2004) Inhibition of Epstein–Barr virus-induced growth proliferation by a nuclear antigen EBNA2-TAT peptide. Proc. Natl. Acad. Sci. U.S.A. **101**, 4625–4630

Feederle R, Shannon-Lowe C, Baldwin G, Delecluse HJ (2005). Defective infectious particles and rare packaged genomes produced by cells carrying terminal-repeat-negative epstein-barr virus. *J Virol* **79**: 7641-7647.

Feinberg AP, Vogelstein B (1983). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* **301:** 89-92.

Fejer G, Koroknai A, Banati F, Györy I, Salamon D, Wolf H, Niller HH, Minarovits J (2008). Latency type-specific distribution of epigenetic marks at the alternative promoters Cp and Qp of Epstein-Barr virus. *J Gen Virol* **89:** 1364-1370.

Fernandez AF, Rosales C, Lopez-Nieva P, Graña O, Ballestar E, Ropero S, Espada J, Melo SA, Lujambio A, Fraga MF, Pino I, Javierre B, Carmona FJ, Acquadro F, Steenbergen RD, Snijders PJ, Meijer CJ, Pineau P, Dejean A, Lloveras B, Capella G, Quer J, Buti M, Esteban JI, Allende H, Rodriguez-Frias F, Castellsague X, Minarovits J, Ponce J, Capello D, Gaidano G, Cigudosa JC, Gomez-Lopez G, Pisano DG, Valencia A, Piris MA, Bosch FX, Cahir-McFarland E, Kieff E, Esteller M (2009). The dynamic DNA methylomes of double-stranded DNA viruses associated with human cancer. *Genome Res* **19:** 438-451.

Filion GJ, Zhenilo S, Salozhin S, Yamada D, Prokhortchouk E, Defossez PA (2006). A family of human zinc finger proteins that bind methylated DNA and repress transcription. *Mol Cell Biol* **26:** 169-181.

Flanagan JM (2007). Host epigenetic modifications by oncogenic viruses. *Br J Cancer* **96**: 183-188.

Floettmann JE, Ward K, Rickinson AB, Rowe M (1996). Cytostatic effect of Epstein-Barr virus latent membrane protein-1 analyzed using tetracycline-regulated expression in B cell lines. *Virology* **223**: 29-40.

Flotho C, Claus R, Batz C, Schneider M, Sandrock I, Ihde S, Plass C, Niemeyer CM, Lübbert M (2009). The DNA methyltransferase inhibitors azacitidine, decitabine and zebularine exert differential effects on cancer gene expression in acute myeloid leukemia cells. *Leukemia* **23:** 1019-1028.

Flower K, Hellen E, Newport MJ, Jones S, Sinclair AJ (2010). Evaluation of a prediction protocol to identify potential targets of epigenetic reprogramming by the cancer associated Epstein Barr virus. *PLoS One* **5:** 9443-9449.

Gal-Yam EN, Egger G, Iniguez L, Holster H, Einarsson S, Zhang X, Lin JC, Liang G, Jones PA, Tanay A (2008). Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line. *Proc Natl Acad Sci U S A* **105**: 12979-12984.

Gaudet F, Talbot D, Leonhardt H, Jaenisch R (1998). A short DNA methyltransferase isoform restores methylation in vivo. *J Biol Chem* **273:** 32725-32729.

Gebhard C, Benner C, Ehrich M, Schwarzfischer L, Schilling E, Klug M, Dietmaier W, Thiede C, Holler E, Andreesen R, Rehli M (2010). General transcription factor binding at CpG islands in normal cells correlates with resistance to de novo DNA methylation in cancer cells. *Cancer Res* **70:** 1398-1407.

Gires O, Kohlhuber F, Kilger E, Baumann M, Kieser A, Kaiser C, Zeidler R, Scheffer B, Ueffing M, Hammerschmidt W (1999). Latent membrane protein 1 of Epstein-Barr virus interacts with JAK3 and activates STAT proteins. *EMBO J* **18:** 3064-3073.

Given D, Kieff E. DNA of Epstein-Barr virus (1979). VI. Mapping of the internal tandem reiteration. *J Virol* **31:** 315-324.

Gompels UA, Nicholas J, Lawrence G, Jones M, Thomson BJ, Martin ME, Efstathiou S, Craxton M, Macaulay HA(1995). The DNA sequence of human herpesvirus-6: structure, coding content, and genome evolution. *Virology* **209:**29-51.

Gross H, Barth S, Palermo RD, Mamiani A, Hennard C, Zimber-Strobl U, West MJ, Kremmer E, Grässer FA (2010). Asymmetric Arginine dimethylation of Epstein-Barr virus nuclear antigen 2 promotes DNA targeting. *Virology* **397:** 299-310.

Grundhoff A, Sullivan CS, Ganem D (2006). A combined computational and microarray-based approach identifies novel microRNAs encoded by human gamma-herpesviruses. *RNA* **12:** 733-750.

Gutensohn N, Cole P (1980). Epidemiology of Hodgkin's disease. *Semin Oncol* **7:** 92-102.

Günther T, Grundhoff A (2010). The epigenetic landscape of latent Kaposi sarcoma-associated herpesvirus genomes. *PLoS Pathog* e1000935.

Hansen RS, Wijmenga C, Luo P, Stanek AM, Canfield TK, Weemaes CM, Gartler SM (1999). The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. *Proc Natl Acad Sci U S A* **96:** 14412-14417.

Hata K, Okano M, Lei H, Li E (2002). Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development* **129:** 1983-1993.

Hayes DP, Brink AATP, Vervoort MB, Middeldorp JM, Meijer CJ, van den Brule AJ (1999). Expression of Epstein-Barr virus (EBV) transcripts encoding homologues to important human proteins in diverse EBV associated diseases. *Mol Pathol* **52:** 97-103.

Heather J, Flower K, Isaac S, Sinclair AJ (2009). The Epstein-Barr virus lytic cycle activator Zta interacts with methylated ZRE in the promoter of host target gene egr1. *J Gen Virol* **90:** 1450-1454.

Heid CA, Stevens J, Livak KJ, Williams PM (1996). Real time quantitative PCR. *Genome Res* **6**: 986-994.

Henderson A, Ripley S, Heller M, Kieff E (1983). Chromosome site for Epstein-Barr virus DNA in a Burkitt tumor cell line and in lymphocytes growth-transformed in vitro. *Proc Natl Acad Sci U S A* **80:** 1987-1991.

Hendrich B, Guy J, Ramsahoye B, Wilson VA, Bird A (2001). Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes Dev* **15:** 710-723.

Henle W, Diehl V, Kohn G, Zur Hausen H, Henle G (1967). Herpes-type virus and chromosome marker in normal leukocytes after growth with irradiated Burkitt cells. *Science* **157:** 1064-1065.

Herman JG, Baylin SB (2003). Gene silencing in cancer in association with promoter hypermethylation. *N Engl J Med* **349:** 2042-2454.

Hino R, Uozaki H, Murakami N, Ushiku T, Shinozaki A, Ishikawa S, Morikawa T, Nakaya T, Sakatani T, Takada K, Fukayama M (2009). Activation of DNA methyltransferase 1 by EBV latent membrane protein 2A leads to promoter hypermethylation of PTEN gene in gastric carcinoma. *Cancer Res* **69:** 2766-2764.

Hino R, Uozaki H, Murakami N, Ushiku T, Shinozaki A, Ishikawa S, Morikawa T, Nakaya T, Sakatani T, Takada K, Fukayama M (2009). Activation of DNA methyltransferase 1 by EBV latent membrane protein 2A leads to promoter hypermethylation of PTEN gene in gastric carcinoma. *Cancer Res* **69:** 2766-2774.

Hollyoake M, Stühler A, Farrell P, Gordon J, Sinclair A (1995). The normal cell cycle activation program is exploited during the infection of quiescent B lymphocytes by Epstein-Barr virus. *Cancer Res* **55:** 4784-4787.

Honda T, Tamura G, Waki T, Kawata S, Terashima M, Nishizuka S, Motoyama T (2004). Demethylation of MAGE promoters during gastric cancer progression. *Br J Cancer* **90:** 838-843.

Hu J, Fan H, Liu D, Zhang S, Zhang F, Xu H (2010). DNMT3B promoter polymorphism and risk of gastric cancer. *Dig Dis Sci* **55:** 1011-1016.

Huen DS, Henderson SA, Croom-Carter D, Rowe M (1995). The Epstein-Barr virus latent membrane protein-1 (LMP1) mediates activation of NF-kappa B and cell surface phenotype via two effector regions in its carboxy-terminal cytoplasmic domain. *Oncogene* **10:** 549-560.

Hutchings IA, Tierney RJ, Kelly GL, Stylianou J, Rickinson AB, Bell AI (2006). Methylation status of the Epstein-Barr virus (EBV) BamHI W latent cycle promoter and promoter activity: analysis with novel EBV-positive Burkitt and lymphoblastoid cell lines. *J Virol* **80**: 10700-10711.

Illingworth RS, Bird AP (2009). CpG islands--'a rough guide'. *FEBS Lett* **583:** 1713-1720.

Imai S, Koizumi S, Sugiura M, Tokunaga M, Uemura Y, Yamamoto N, Tanaka S, Sato E, Osato T (1994). Gastric carcinoma: monoclonal epithelial malignant cells expressing Epstein-Barr virus latent infection protein. *Proc Natl Acad Sci U S A* **91:** 9131-9135.

Infantino S, Benz B, Waldmann T, Jung M, Schneider R, Reth M (2010). Arginine methylation of the B cell antigen receptor promotes differentiation. *J Exp Med* **207**: 711-719.

Jair KW, Bachman KE, Suzuki H, Ting AH, Rhee I, Yen RW, Baylin SB, Schuebel KE (2006). De novo CpG island methylation in human cancer cells. *Cancer Res* **66:** 682-692.

Jansson A, Masucci M, Rymo L (1992). Methylation of discrete sites within the enhancer region regulates the activity of the Epstein-Barr virus BamHI W promoter in Burkitt lymphoma lines. *J Virol* **66:** 62-69.

Jeltsch A (2006). Molecular enzymology of mammalian DNA methyltransferases. *Curr Top Microbiol Immunol* **301:** 203-225.

Jin B, Tao Q, Peng J, Soo HM, Wu W, Ying J, Fields CR, Delmas AL, Liu X, Qiu J, Robertson KD (2008). DNA methyltransferase 3B (DNMT3B) mutations in ICF syndrome lead to altered epigenetic modifications and aberrant expression of genes regulating development, neurogenesis and immune function. *Hum Mol Genet* **17:** 690-709.

Jin SG, Guo C, Pfeifer GP (2008). GADD45A does not promote DNA demethylation. *PLoS Genet* **4:** 1000013-1000022.

Jones PA, Laird PW (1999). Cancer epigenetics comes of age. *Nat Genet* **21:** 163-167.

Jones PA, Liang G (2009). Rethinking how DNA methylation patterns are maintained. *Nat Rev Genet* **10:** 805-811.

Kalla M, Schmeinck A, Bergbauer M, Pich D, Hammerschmidt W (2010). AP-1 homolog BZLF1 of Epstein-Barr virus has two essential functions dependent on the epigenetic state of the viral genome. *Proc Natl Acad Sci U S A* **107:** 850-855.

Kamesaki H, Fukuhara S, Tatsumi E, Uchino H, Yamabe H, Miwa H, Shirakawa S, Hatanaka M, Honjo T (1986). Cytochemical, immunologic, chromosomal, and molecular genetic analysis of a novel cell line derived from Hodgkin's disease. *Blood* **68**: 285-292.

Kanai Y, Ushijima S, Kondo Y, Nakanishi Y, Hirohashi S (2001). DNA methyltransferase expression and DNA methylation of CPG islands and peri-centromeric satellite regions in human colorectal and stomach cancers. *Int J Cancer* **91:** 205-512.

Kangaspeska S, Stride B, Métivier R, Polycarpou-Schwarz M, Ibberson D, Carmouche RP, Benes V, Gannon F, Reid G (2008). Transient cyclical methylation of promoter DNA. *Nature* **452:** 112-115.

Kantarjian HM, O'Brien S, Cortes J, Giles FJ, Faderl S, Issa JP, Garcia-Manero G, Rios MB, Shan J, Andreeff M, Keating M, Talpaz M (2003). Results of decitabine (5-aza-2'deoxycytidine) therapy in 130 patients with chronic myelogenous leukemia. *Cancer* **98:** 522-528.

Kanzler H, Hansmann ML, Kapp U, Wolf J, Diehl V, Rajewsky K, Kuppers R (1996b). Molecular single cell analysis demonstrates the derivation of a peripheral blood-derived cell line (L1236) from the Hodgkin/Reed-Sternberg cells of a Hodgkin's lymphoma patient. *Blood* **87**: 3429-3436.

Kareta MS, Botello ZM, Ennis JJ, Chou C, Chédin F (2006). Reconstitution and mechanism of the stimulation of de novo methylation by human DNMT3L. *J Biol Chem* **281:** 25893-25902.

Karlsson QH, Schelcher C, Verrall E, Petosa C, Sinclair AJ (2008). Methylated DNA recognition during the reversal of epigenetic silencing is regulated by cysteine and serine residues in the Epstein-Barr virus lytic switch protein. *PLoS Pathog* **4:** 1000005-100011.

Kass SU, Pruss D, Wolffe AP (1997). How does DNA methylation repress transcription? *Trends Genet* **13:** 444-449.

Kaye KM, Izumi KM, Kieff E (1993). Epstein-Barr virus latent membrane protein 1 is essential for B-lymphocyte growth transformation. *Proc Natl Acad Sci U S A* **90:**9150-4.

Kieff E and Rickinson AB. In: *Fields Virology*. Eds: Knipe DM, Howley PM, Lippincott Williams and Wilkins, Philadelphia, 2007

Kieser A, Kilger E, Gires O, Ueffing M, Kolch W, Hammerschmidt W (1997). Epstein-Barr virus latent membrane protein-1 triggers AP-1 activity via the c-Jun N-terminal kinase cascade. *EMBO J* **16:** 6478-6485.

Kim KH, Choi JS, Kim IJ, Ku JL, Park JG (2006). Promoter hypomethylation and reactivation of MAGE-A1 and MAGE-A3 genes in colorectal cancer cell lines and cancer tissues. *World J Gastroenterol* **12:** 5651-5657.

Kintner CR, Sugden B (1979). The structure of the termini of the DNA of Epstein-Barr virus. *Cell* **17:** 661-671.

Klein G (1983). Specific chromosomal translocations and the genesis of B-cell-derived tumors in mice and men. *Cell* **32:** 311-315.

Kleinschmidt MA, Streubel G, Samans B, Krause M, Bauer UM (2008). The protein arginine methyltransferases CARM1 and PRMT1 cooperate in gene regulation. *Nucleic Acids Res* **36:** 3202-3213.

Klimasauskas S, Kumar S, Roberts RJ, Cheng X (1994). HhaI methyltransferase flips its target base out of the DNA helix. *Cell* **76:** 357-369.

Klose RJ, Bird AP (2006). Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* **31:** 89-97.

Koga Y, Pelizzola M, Cheng E, Krauthammer M, Sznol M, Ariyan S, Narayan D, Molinaro AM, Halaban R, Weissman SM (2009). Genome-wide screen of promoter methylation identifies novel markers in melanoma. *Genome Res* **19:** 1462-1470.

Kouzarides T (2007). Chromatin modifications and their function. *Cell* **128:** 693-705.

Krause CD, Yang ZH, Kim YS, Lee JH, Cook JR, Pestka S (2007). Protein arginine methyltransferases: evolution and assessment of their pharmacological and therapeutic potential. *Pharmacol Ther* **113:** 50-87.

Kudoh A, Fujita M, Kiyono T, Kuzushima K, Sugaya Y, Izuta S, Nishiyama Y, Tsurumi T (2003). Reactivation of lytic replication from B cells latently infected with Epstein-Barr virus occurs with high S-phase cyclin-dependent kinase activity while inhibiting cellular DNA replication. *J Virol* **77:** 851-861.

Küppers R (2009). Molecular biology of Hodgkin lymphoma. *Hematology Am Soc Hematol Educ Program* 491-496.

Lacroix M, El Messaoudi S, Rodier G, Le Cam A, Sardet C, Fabbrizio E (2008). The histone-binding protein COPR5 is required for nuclear functions of the protein arginine methyltransferase PRMT5. *EMBO Rep* **9:** 452-458.

Laird PW (2005). Cancer epigenetics. *Hum Mol Genet* **14:** 65-76.

Landais E, Saulquin X, Houssaint E (2005). The human T cell immune response to Epstein-Barr virus. *Int J Dev Biol* **49:** 285-292.

Lee MA, Diamond ME, Yates JL (1999). Genetic evidence that EBNA-1 is needed for efficient, stable latent infection by Epstein-Barr virus. *J Virol* **73**: 2974-2982.

Lee J, Bedford MT (2002). PABP1 identified as an arginine methyltransferase substrate using high-density protein arrays. *EMBO Rep* **3:** 268-273.

Levine PH, Ablashi DV, Berard CW, Carbone PP, Waggoner DE, Malan L (1971). Elevated antibody titers to Epstein-Barr virus in Hodgkin's disease. *Cancer* **27:** 416-421.

Li E, Bestor TH, Jaenisch R (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69:** 915-926.

Liang G, Lin JC, Wei V, Yoo C, Cheng JC, Nguyen CT, Weisenberger DJ, Egger G, Takai D, Gonzales FA, Jones PA (2004). Distinct localization of histone H3 acetylation and H3-K4 methylation to the transcription start sites in the human genome. *Proc Natl Acad Sci U S A* **101:** 7357-7362.

Lin TS, Lee H, Chen RA, Ho ML, Lin CY, Chen YH, Tsai YY, Chou MC, Cheng YW (2005). An association of DNMT3b protein expression with P16INK4a promoter hypermethylation in non-smoking female lung cancer with human papillomavirus infection. *Cancer Lett* **226:** 77-84.

Livak KJ, Schmittgen TD (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**: 402-408.

Longnecker R (2000). Epstein-Barr virus latency: LMP2, a regulator or means for Epstein-Barr virus persistence? *Adv Cancer Res* **79:** 175-200.

Longnecker R, Kieff E (1990). A second Epstein-Barr virus membrane protein (LMP2) is expressed in latent infection and colocalizes with LMP1. *J Virol* **64:** 2319-2326.

Longnecker R, Miller CL (1996). Regulation of Epstein-Barr virus latency by latent membrane protein 2. *Trends Microbiol* **4:** 38-42.

Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389:** 251-260.

Magrath I (1990). The pathogenesis of Burkitt's lymphoma. *Adv Cancer Res* **55:**133-270.

Manolov G, Manolova Y (1972). Marker band in one chromosome 14 from Burkitt lymphomas. *Nature* **237:** 33-34.

Martin-Subero JI, Ammerpohl O, Bibikova M, Wickham-Garcia E, Agirre X, Alvarez S, Brüggemann M, Bug S, Calasanz MJ, Deckert M, Dreyling M, Du MQ, Dürig J, Dyer MJ, Fan JB, Gesk S, Hansmann ML, Harder L, Hartmann S, Klapper W, Küppers R, Montesinos-Rongen M, Nagel I, Pott C, Richter J, Román-Gómez J, Seifert M, Stein H, Suela J, Trümper L, Vater I, Prosper F, Haferlach C, Cruz Cigudosa J, Siebert R (2009). A comprehensive microarray-based DNA methylation study of 367 hematological neoplasms. *PLoS One* **4:** 6986-6995.

Masucci MG, Contreras-Salazar B, Ragnar E, Falk K, Minarovits J, Ernberg I, Klein G (1989). 5-Azacytidine up regulates the expression of Epstein-Barr virus nuclear antigen 2 (EBNA-2) through EBNA-6 and latent membrane protein in the Burkitt's lymphoma line rael. *J Virol* **63:** 3135-3141.

Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, Johnson BE, Hong C, Nielsen C, Zhao Y, Turecki G, Delaney A, Varhol R, Thiessen N, Shchors K, Heine VM, Rowitch DH, Xing X, Fiore C, Schillebeeckx M, Jones SJ, Haussler D, Marra MA, Hirst M, Wang T, Costello JF (2010). Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **466**: 253-257.

Métivier R, Gallais R, Tiffoche C, Le Péron C, Jurkowska RZ, Carmouche RP, Ibberson D, Barath P, Demay F, Reid G, Benes V, Jeltsch A, Gannon F, Salbert G (2010). Cyclical DNA methylation of a transcriptionally active promoter. *Nature* **452:** 45-50.

Miller G, Lipman M (1973). Release of infectious Epstein-Barr virus by transformed marmoset leukocytes. *Proc Natl Acad Sci U S A* **70:** 190-194.

Miller G, Rabson M, Heston L (1984). Epstein-Barr virus with heterogeneous DNA disrupts latency. *J Virol* **50**: 174-182.

Minarovits J, Hu LF, Minarovits-Kormuta S, Klein G, Ernberg I (1994). Sequence-specific methylation inhibits the activity of the Epstein-Barr virus LMP 1 and BCR2 enhancer-promoter regions. *Virology* **200:** 661-667.

Minarovits J, Hu LF, Minarovits-Kormuta S, Klein G, Ernberg I (1994). Sequence-specific methylation inhibits the activity of the Epstein-Barr virus LMP 1 and BCR2 enhancer-promoter regions. *Virology* **200:** 661-667.

Moghaddam A, Rosenzweig M, Lee-Parritz D, Annis B, Johnson RP, Wang F (1997). An animal model for acute and persistent Epstein-Barr virus infection. *Science* **276:** 2030-2033.

Molesworth SJ, Lake CM, Borza CM, Turk SM, Hutt-Fletcher LM (2000). Epstein-Barr virus gH is essential for penetration of B cells but also plays a role in attachment of virus to epithelial cells. *J Virol* **74:** 6324-6332.

Morrow RH (1985). Epidemiological evidence for the role of falciparum malaria in the pathogenesis of Burkitt's lymphoma. *IARC Sci Publ***. 60:** 177-186.

Murphy G, Pfeiffer R, Camargo MC, Rabkin CS (2009). Meta-analysis shows that prevalence of Epstein-Barr virus-positive gastric cancer differs based on sex and anatomic location. *Gastroenterology* **137:** 824-833.

Nan X, Meehan RR, Bird A (1993). Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Res* **21:** 4886-4892.

Navarro A, Gaya A, Martinez A, Urbano-Ispizua A, Pons A, Balagué O, Gel B, Abrisqueta P, Lopez-Guillermo A, Artells R, Montserrat E, Monzo M (2008). MicroRNA expression profiling in classic Hodgkin lymphoma. *Blood* **111**: 2825-2832.

Niller HH, Wolf H, Minarovits J (2008). Regulation and dysregulation of Epstein-Barr virus latency: implications for the development of autoimmune diseases. *Autoimmunity* **41:** 298-328.

Ohm JE, McGarvey KM, Yu X, Cheng L, Schuebel KE, Cope L, Mohammad HP, Chen W, Daniel VC, Yu W, Berman DM, Jenuwein T, Pruitt K, Sharkis SJ, Watkins DN, Herman JG, Baylin SB (2007). A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nat Genet* **39:** 237-242.

Okano M, Li E (2002). Genetic analyses of DNA methyltransferase genes in mouse model system. *J Nutr* **132:** 2462-2465.

Pal S, Baiocchi RA, Byrd JC, Grever MR, Jacob ST, Sif S (2007). Low levels of miR-92b/96 induce PRMT5 translation and H3R8/H4R3 methylation in mantle cell lymphoma. *EMBO J* **26**: 3558-3569.

Pal S, Yun R, Datta A, Lacomis L, Erdjument-Bromage H, Kumar J, Tempst P, Sif S (2003). mSin3A/histone deacetylase 2- and PRMT5-containing Brg1 complex is involved in transcriptional repression of the Myc target gene cad. *Mol Cell Biol* **23:** 7475-7487.

Park IY, Sohn BH, Yu E, Suh DJ, Chung YH, Lee JH, Surzycki SJ, Lee YI (2007). Aberrant epigenetic modifications in hepatocarcinogenesis induced by hepatitis B virus X protein. *Gastroenterology* **132:** 1476-1494.

Park JH, Jeon JP, Shim SM, Nam HY, Kim JW, Han BG, Lee S (2007). Wp specific methylation of highly proliferated LCLs. *Biochem Biophys Res Commun* **358:** 513-520.

Paulson EJ, Speck SH (1999). Differential methylation of Epstein-Barr virus latency promoters facilitates viral persistence in healthy seropositive individuals. *J Virol* **73:** 9959-9968.

Pearson GR, Luka J, Petti L, Sample J, Birkenbach M, Braun D, Kieff E (1987). Identification of an Epstein-Barr virus early gene encoding a second component of the restricted early antigen complex. *Virology* **160:** 151-161.

Pisani P, Parkin DM, Muñoz N, Ferlay J (1997). Cancer and infection: estimates of the attributable fraction in 1990. *Cancer Epidemiol Biomarkers Prev* **6:** 387-400.

Pfeffer S, Zavolan M, Grässer FA, Chien M, Russo JJ, Ju J, John B, Enright AJ, Marks D, Sander C, Tuschl T (2004). Identification of virus-encoded microRNAs. *Science* **304:**734-736.

Raab-Traub N, Flynn K (1986). The structure of the termini of the Epstein-Barr virus as a marker of clonal cellular proliferation. *Cell* **47:** 883-889.

Rauch TA, Wu X, Zhong X, Riggs AD, Pfeifer GP (2009). A human B cell methylome at 100-base pair resolution. *Proc Natl Acad Sci U S A* **106:** 671-678.

Reed BD, Charos AE, Szekely AM, Weissman SM, Snyder M (2008). Genome-wide occupancy of SREBP1 and its partners NFY and SP1 reveals novel functional roles and combinatorial regulation of distinct classes of genes. *PLoS Genet* **4**: 1000133-1000140.

Reik W (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447:** 425-432.

Richard S, Morel M, Cléroux P (2005). Arginine methylation regulates IL-2 gene expression: a role for protein arginine methyltransferase 5 (PRMT5). *Biochem J* **388:** 379-386.

Rivailler P, Cho YG, Wang F (2002). Complete genomic sequence of an Epstein-Barr virus-related herpesvirus naturally infecting a new world primate: a defining point in the evolution of oncogenic lymphocryptoviruses. *J Virol* **76:** 12055-12068.

Roberts ML, Cooper NR (1998). Activation of a ras-MAPK-dependent pathway by Epstein-Barr virus latent membrane protein 1 is essential for cellular transformation. *Virology* **240:** 93-99.

Robertson ES, Lin J, Kieff E (1996). The amino-terminal domains of Epstein-Barr virus nuclear proteins 3A, 3B, and 3C interact with RBPJ (kappa). *J Virol* **70:** 3068-3074.

Robertson KD (2001). DNA methylation, methyltransferases, and cancer. *Oncogene* **20:** 139-155.

Robertson KD (2005). DNA methylation and human disease. *Nat Rev Genet* **6**: 597-610.

Robertson KD, Manns A, Swinnen LJ, Zong JC, Gulley ML, Ambinder RF (1996). CpG methylation of the major Epstein-Barr virus latency promoter in Burkitt's lymphoma and Hodgkin's disease. *Blood* **88:** 3129-3136.

Rooney CM, Rowe DT, Ragot T, Farrell PJ (1989). The spliced BZLF1 gene of Epstein-Barr virus (EBV) transactivates an early EBV promoter and induces the virus productive cycle. *J Virol* **63:** 3109-3116.

Roughan JE, Torgbor C, Thorley-Lawson DA (2010). Germinal center B cells latently infected with Epstein-Barr virus proliferate extensively but do not increase in number. *J Virol* **84:** 1158-1168.

Rowe M, Rowe DT, Gregory CD, Young LS, Farrell PJ, Rupani H, Rickinson AB (1987a). Differences in B cell growth phenotype reflect novel patterns of Epstein–Barr virus latent gene expression in Burkitt's lymphoma cells. *EMBO J* **6:** 2743–2751.

Ruike Y, Imanaka Y, Sato F, Shimizu K, Tsujimoto G (2010). Genome-wide analysis of aberrant methylation in human breast cancer cells using methyl-DNA immunoprecipitation combined with high-throughput sequencing. *BMC Genomics* **11**: 137-145.

Saito Y, Kanai Y, Nakagawa T, Sakamoto M, Saito H, Ishii H, Hirohashi S (2003). Increased protein expression of DNA methyltransferase (DNMT) 1 is significantly correlated with the malignant potential and poor prognosis of human hepatocellular carcinomas. *Int J Cancer* **105:** 527-532.

Sakatani T, Kaneda A, Iacobuzio-Donahue CA, Carter MG, de Boom Witzel S, Okano H, Ko MS, Ohlsson R, Longo DL, Feinberg AP (2005). Loss of imprinting of Igf2 alters intestinal maturation and tumorigenesis in mice. *Science* **307:** 1976-1978.

Salamon D, Takacs M, Ujvari D, Uhlig J, Wolf H, Minarovits J, Niller HH (2001). Protein-DNA binding and CpG methylation at nucleotide resolution of latency-associated promoters Qp, Cp, and LMP1p of Epstein-Barr virus. *J Virol* **75:** 2584-2596.

Sample J, Young L, Martin B, Chatman T, Kieff E, Rickinson A, Kieff E (1990). Epstein-Barr virus types 1 and 2 differ in their EBNA-3A, EBNA-3B, and EBNA-3C genes. *J Virol* **64:** 4084-4092.

Santos F, Hendrich B, Reik W, Dean W (2002). Dynamic reprogramming of DNA methylation in the early mouse embryo. *Dev Biol* **241:** 172-182.

Sasai N, Defossez PA (2009). Many paths to one goal? The proteins that recognize methylated DNA in eukaryotes. *Int J Dev Biol* **53:** 323-334.

Sawada M, Kanai Y, Arai E, Ushijima S, Ojima H, Hirohashi S (2007). Increased expression of DNA methyltransferase 1 (DNMT1) protein in uterine cervix squamous cell carcinoma and its precursor lesion. *Cancer Lett* **251:** 211-219.

Schaadt M, Diehl V, Stein H, Fonatsch C, Kirchner HH (1980). Two neoplastic cell lines with unique features derived from Hodgkin's disease. *Int J Cancer* **26**: 723-731.

Schaefer M, Pollex T, Hanna K, Tuorto F, Meusburger M, Helm M, Lyko F (2010). RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes Dev* **24:** 1590-1595.

Schlager S, Speck SH, Woisetschläger M (1996). Transcription of the Epstein-Barr virus nuclear antigen 1 (EBNA1) gene occurs before induction of the BCR2 (Cp) EBNA gene promoter during the initial stages of infection in B cells. *J Virol* **70:** 3561-3570.

Schlesinger Y, Straussman R, Keshet I, Farkash S, Hecht M, Zimmerman J, Eden E, Yakhini Z, Ben-Shushan E, Reubinoff BE, Bergman Y, Simon I, Cedar H (2007). Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat Genet* **39:** 232-236.

Schmitz R, Stanelle J, Hansmann ML, Küppers R (2009). Pathogenesis of classical and lymphocyte-predominant Hodgkin lymphoma. *Annu Rev Pathol* **4:**151-174.

Sen GL, Reuter JA, Webster DE, Zhu L, Khavari PA (2010). DNMT1 maintains progenitor function in self-renewing somatic tissue. *Nature* **463:** 563-567.

Seo SY, Kim EO, Jang KL (2008). Epstein-Barr virus latent membrane protein 1 suppresses the growth-inhibitory effect of retinoic acid by inhibiting retinoic acid receptor-beta2 expression via DNA methylation. *Cancer Lett* **270:** 66-76.

Shah KM, Young LS (2009). Epstein-Barr virus and carcinogenesis: beyond Burkitt's lymphoma. *Clin Microbiol Infect* **15**: 982-988.

Shamay M, Krithivas A, Zhang J, Hayward SD (2006). Recruitment of the de novo DNA methyltransferase Dnmt3a by Kaposi's sarcoma-associated herpesvirus LANA. *Proc Natl Acad Sci U S A* **103:** 14554-14559.

Shann YJ, Cheng C, Chiao CH, Chen DT, Li PH, Hsu MT (2008). Genome-wide mapping and characterization of hypomethylated sites in human tissues and breast cancer cell lines. *Genome Res* **18**: 791-801.

Shannon-Lowe C, Baldwin G, Feederle R, Bell A, Rickinson A, Delecluse HJ (2005). Epstein-Barr virus-induced B-cell transformation: quantitating events from virus binding to cell outgrowth. *J Gen Virol* **86:** 3009-3019.

Sharma S, Kelly TK, Jones PA (2010). Epigenetics in cancer. *Carcinogenesis* **31:** 27-36.

Shire K, Kapoor P, Jiang K, Hing MN, Sivachandran N, Nguyen T, Frappier L (2006). Regulation of the EBNA1 Epstein-Barr virus protein by serine phosphorylation and arginine methylation. *J Virol* **80:** 5261-5272.

Siemer D, Kurth J, Lang S, Lehnerdt G, Stanelle J, Küppers R (2008). EBV transformation overrides gene expression patterns of B cell differentiation stages. *Mol Immunol* **45:** 3133-3141.

Singh KP, Kumari R, Pevey C, Jackson D, DuMond JW (2009). Long duration exposure to cadmium leads to increased cell survival, decreased DNA repair capacity, and genomic instability in mouse testicular Leydig cells. *Cancer Lett* **279:** 84-92.

Skare J, Farley J, Strominger JL, Fresen KO, Cho MS, zur Hausen H (1985). Transformation by Epstein-Barr virus requires DNA sequences in the region of BamHI fragments Y and H. *J Virol* **55**: 286-297.

Suetake I, Shinozaki F, Miyagawa J, Takeshima H, Tajima S (2004). DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. *J Biol Chem* **279:** 27816-27823.

Szyf M, Eliasson L, Mann V, Klein G, Razin A (1985). Cellular and viral DNA hypomethylation associated with induction of Epstein-Barr virus lytic cycle. *Proc Natl Acad Sci U S A* **82:**8090-80944.

Takacs M, Banati F, Koroknai A, Segesdi J, Salamon D, Wolf H, Niller HH, Minarovits J (2010). Epigenetic regulation of latent Epstein-Barr virus promoters. *Biochim Biophys Acta* **1799:** 228-235.

Takai D, Jones PA (2002). Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci U S A* **99:** 3740-3745.

Tang J, Frankel A, Cook RJ, Kim S, Paik WK, Williams KR, Clarke S, Herschman HR (2000). PRMT1 is the predominant type I protein arginine methyltransferase in mammalian cells. *J Biol Chem* **275:** 7723-7730.

Taniguchi Y, Nosaka K, Yasunaga J, Maeda M, Mueller N, Okayama A, Matsuoka M (2005). Silencing of human T-cell leukemia virus type I gene transcription by epigenetic mechanisms. *Retrovirology* **2:** 64-72.

Tao Q, Robertson KD (2003). Stealth technology: how Epstein-Barr virus utilizes DNA methylation to cloak itself from immune detection. *Clin Immunol* **109:** 53-63.

Tao Q, Robertson KD, Manns A, Hildesheim A, Ambinder RF (1998). The Epstein-Barr virus major latent promoter Qp is constitutively active, hypomethylated, and methylation sensitive. *J Virol* **72:** 7075-7083.

Taylor KH, Rahmatpanah F, Davis JW, Caldwell CW (2008). Chromosomal localization of DNA methylation in small B-cell lymphoma. *Leukemia* **22**: 638-641.

Thorley-Lawson DA (2001). Epstein-Barr virus: exploiting the immune system. *Nat Rev Immunol* **1:** 75-82.

Thorley-Lawson DA, Allday MJ (2008). The curious case of the tumour virus: 50 years of Burkitt's lymphoma. *Nat Rev Microbiol* **6:** 913-924.

Tierney RJ, Kirby HE, Nagra JK, Desmond J, Bell AI, Rickinson AB (2000). Methylation of transcription factor binding sites in the Epstein-Barr virus latent cycle promoter Wp coincides with promoter down-regulation during virus-induced B-cell transformation. *J Virol* **74:** 10468-10479.

Tomkinson B, Kieff E (1992). Use of second-site homologous recombination to demonstrate that Epstein-Barr virus nuclear protein 3B is not important for lymphocyte infection or growth transformation in vitro. *J Virol* **66**: 2893-2903.

Tomkinson B, Robertson E, Kieff E (1993). Epstein-Barr virus nuclear proteins EBNA-3A and EBNA-3C are essential for B-lymphocyte growth transformation. *J Virol* **67**: 2014-2025.

Tsai CL, Li HP, Lu YJ, Hsueh C, Liang Y, Chen CL, Tsao SW, Tse KP, Yu JS, Chang YS (2006). Activation of DNA methyltransferase 1 by EBV LMP1 Involves c-Jun NH(2)-terminal kinase signaling. *Cancer Res* **66:** 11668-11676.

Tsai CN, Tsai CL, Tse KP, Chang HY, Chang YS (2002). The Epstein-Barr virus oncogene product, latent membrane protein 1, induces the downregulation of E-cadherin gene

expression via activation of DNA methyltransferases. *Proc Natl Acad Sci U S A* **99:** 10084-10089.

Turek-Plewa J, Jagodziński PP (2005). The role of mammalian DNA methyltransferases in the regulation of gene expression. *Cell Mol Biol Lett* **10:** 631-647.

Ushmorov A, Leithäuser F, Sakk O, Weinhaüsel A, Popov SW, Möller P, Wirth T (2006). Epigenetic processes play a major role in B-cell-specific gene silencing in classical Hodgkin lymphoma. *Blood* **107:** 2493-2500.

Ushmorov A, Ritz O, Hummel M, Leithäuser F, Möller P, Stein H, Wirth T (2004). Epigenetic silencing of the immunoglobulin heavy-chain gene in classical Hodgkin lymphoma-derived cell lines contributes to the loss of immunoglobulin expression. *Blood* **104:** 3326-3334.

van Beek J, Brink AA, Vervoort MB, van Zijp MJ, Meijer CJ, van den Brule AJ, Middeldorp JM (2003). In vivo transcription of the Epstein-Barr virus (EBV) BamHI-A region without associated in vivo BARF0 protein expression in multiple EBV-associated disorders. *J Gen Virol* **84:** 2647-2659.

Vockerodt M, Morgan SL, Kuo M, Wei W, Chukwuma MB, Arrand JR, Kube D, Gordon J, Young LS, Woodman CB, Murray PG (2008). The Epstein-Barr virus oncoprotein, latent membrane protein-1, reprograms germinal centre B cells towards a Hodgkin's Reed-Sternberg-like phenotype. *J Pathol* **216**: 83-92.

Wade PA (2001). Methyl CpG-binding proteins and transcriptional repression. *Bioessays* **23:**1131-1137.

Wang F, Gregory C, Sample C, Rowe M, Liebowitz D, Murray R, Rickinson A, Kieff E (1990). Epstein–Barr virus latent membrane protein (LMP1) and nuclear proteins 2 and 3C are effectors of phenotypic changes in B lymphocytes: EBNA-2 and LMP1 cooperatively induce CD23. *J Virol* **64:** 2309–2318.

Wang J, Sugden B (2005). Origins of bidirectional replication of Epstein-Barr virus: models for understanding mammalian origins of DNA synthesis. *J Cell Biochem* **94:** 247-256.

Wang L, Pal S, Sif S (2008). Protein arginine methyltransferase 5 suppresses the transcription of the RB family of tumor suppressors in leukemia and lymphoma cells. *Mol Cell Biol* **28:** 6262-6277.

Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, Schübeler D (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* **37:** 853-862.

Weber M, Hellmann I, Stadler MB, Ramos L, Pääbo S, Rebhan M, Schübeler D (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet* **39:** 457-66.

Weir JP (1998). Genomic organization and evolution of the human herpesviruses. *Virus Genes* **16**:85-93.

Weiss LM, Movahed LA (1989). In situ demonstration of Epstein-Barr viral genomes in viral-associated B cell lymphoproliferations. *Am J Pathol* **134:** 651-659.

Widschwendter M, Fiegl H, Egle D, Mueller-Holzner E, Spizzo G, Marth C, Weisenberger DJ, Campan M, Young J, Jacobs I, Laird PW (2007). Epigenetic stem cell signature in cancer. *Nat Genet* **39:**157-158.

Wolf J, Kapp U, Bohlen H, Kornacker M, Schoch C, Stahl B, Mucke S, von Kalle C, Fonatsch C, Schaefer HE, Hansmann ML, Diehl V (1996). Peripheral blood mononuclear cells of a patient with advanced Hodgkin's lymphoma give rise to permanently growing Hodgkin-Reed Sternberg cells. *Blood* **87**: 3418-3428.

Wong KM, Levine AJ (1986). Identification and mapping of Epstein-Barr virus early antigens and demonstration of a viral gene activator that functions in trans. *J Virol* **60:** 149-156.

Wu X, Gong Y, Yue J, Qiang B, Yuan J, Peng X (2008). Cooperation between EZH2, NSPc1-mediated histone H2A ubiquitination and Dnmt1 in HOX gene silencing. *Nucleic Acids Res* **36:** 3590-3599.

Yajima M, Kanda T, Takada K (2005). Critical role of Epstein-Barr Virus (EBV)-encoded RNA in efficient EBV-induced B-lymphocyte growth transformation. *J Virol* **79:** 4298-4307.

Yang HH, Hu N, Wang C, Ding T, Dunn BK, Goldstein AM, Taylor PR, Lee MP (2010). Influence of genetic background and tissue types on global DNA methylation patterns. *PLoS One* **5**: 9355-9363.

Young LS, Dawson CW, Clark D, Rupani H, Busson P, Tursz T, Johnson A, Rickinson AB (1988). Epstein-Barr virus gene expression in nasopharyngeal carcinoma. *J Gen Virol* **69:**1051-1065.

Young LS, Murray PG (2003). Epstein-Barr virus and oncogenesis: from latent genes to tumours. *Oncogene* **22:** 5108-5121.

Young LS, Rickinson AB (2004). Epstein-Barr virus: 40 years on. Nat Rev Cancer 4: 757-768.

Yun Zhu J, Pfuhl T, Motsch N, Barth S, Nicholls J, Grässer F, Meister G (2009). Identification of Novel Epstein-Barr Virus MicroRNA Genes from Nasopharyngeal Carcinomas. *J Virol **83:** 3333-3341.*

Zhang X, Cheng X (2003). Structure of the predominant protein arginine methyltransferase PRMT1 and analysis of its binding to substrate peptides. *Structure* **11:** 509-520.

Zhang Y, Ng HH, Erdjument-Bromage H, Tempst P, Bird A, Reinberg D (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev* **13:** 1924-1935.

Zhao Q, Rank G, Tan YT, Li H, Moritz RL, Simpson RJ, Cerruti L, Curtis DJ, Patel DJ, Allis CD, Cunningham JM, Jane SM (2009). PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. *Nat Struct Mol Biol* **16:** 304-11.

Zhao X, Jankovic V, Gural A, Huang G, Pardanani A, Menendez S, Zhang J, Dunne R, Xiao A, Erdjument-Bromage H, Allis CD, Tempst P, Nimer SD (2008). Methylation of RUNX1 by PRMT1 abrogates SIN3A binding and potentiates its transcriptional activity. *Genes Dev* **22:** 640-653.

Zheng X, Pontes O, Zhu J, Miki D, Zhang F, Li WX, Iida K, Kapoor A, Pikaard CS, Zhu JK (2008). ROS3 is an RNA-binding protein required for DNA demethylation in Arabidopsis. *Nature* **455:** 1259-1262.

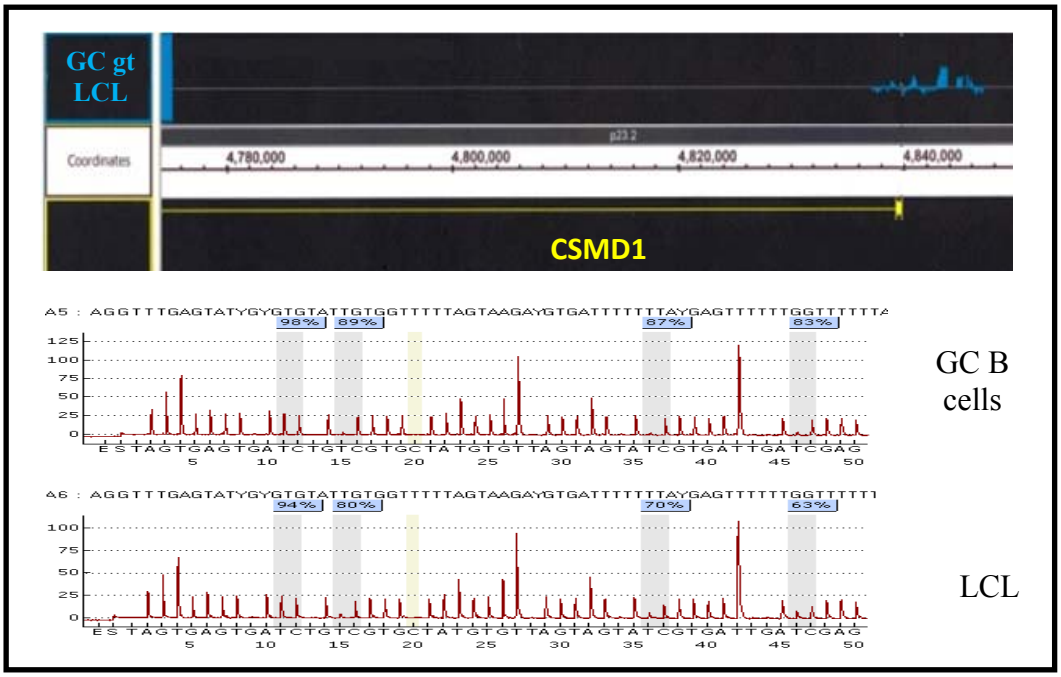
Zilberman D, Coleman-Derr D, Ballinger T, Henikoff S (2008). Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature* **456:** 125-129.

Zinzen RP, Girardot C, Gagneur J, Braun M, Furlong EE (2009). Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature* **462**: 65-70.

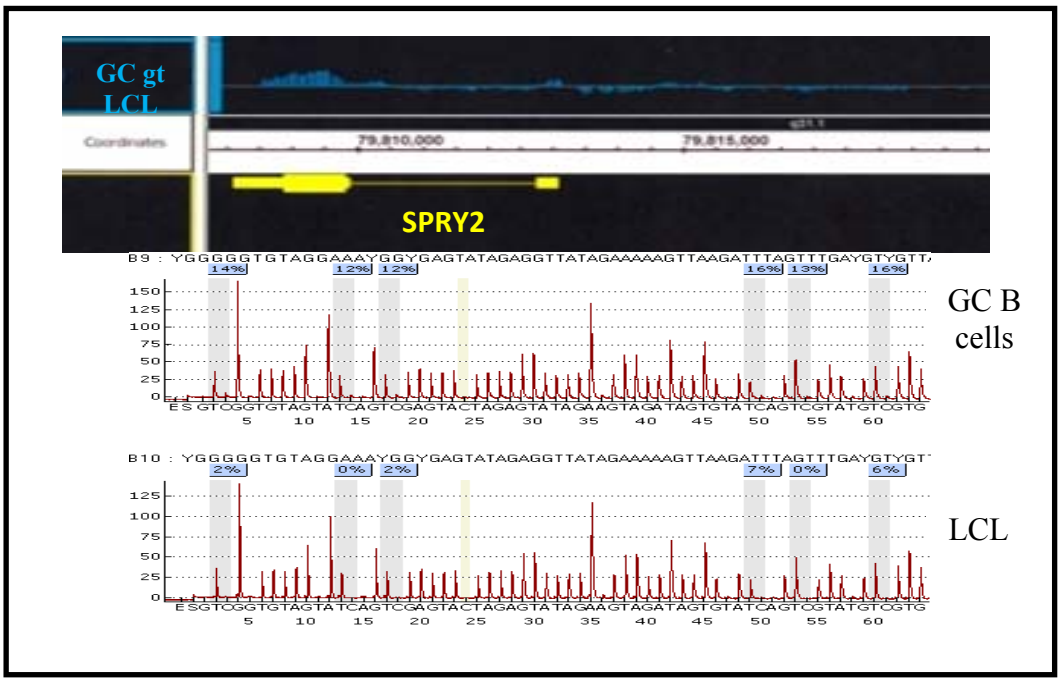
# Annex

Annex 1: Predicted methylation changes confirmed using pyrosequencing. For each of the ten genes shown below, the integrated genome browser was used to visualise the methylation array results (top panel) and pyrosequencing (bottom panel) used to confirm these changes.

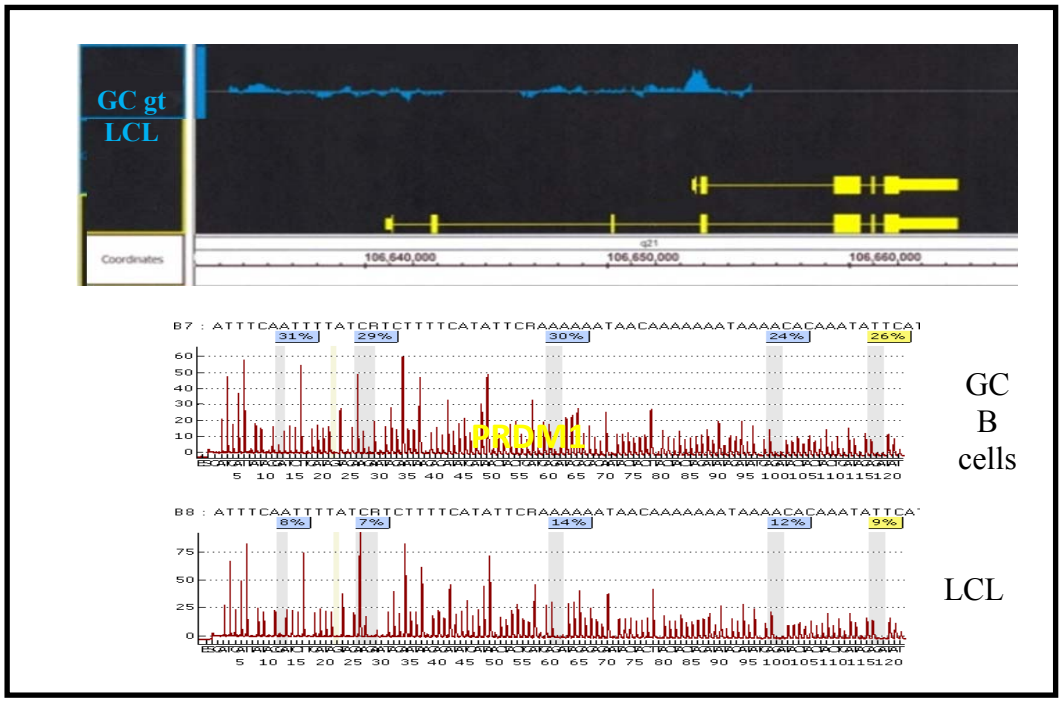
## CSMD1




## SPRY2



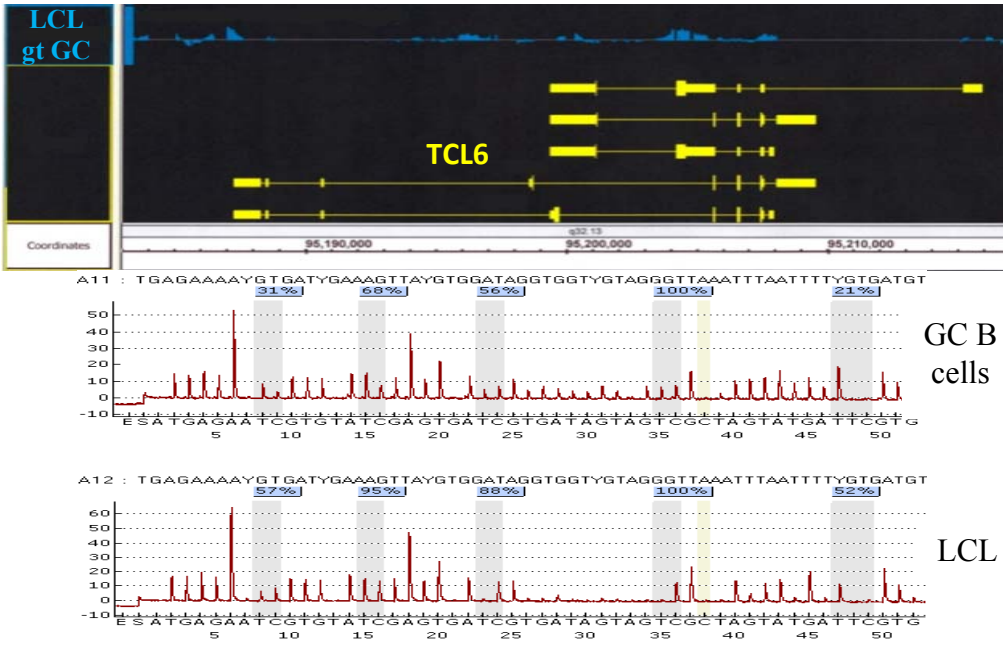
# PRDM1


GC gt
LCL
Coordinates
106,640,000
106,650,000
106,660,000
PRDM1
GC
B
cells
LCL

# GRB10


GC gt
LCL
Coordinates
50,740,000
50,760,000
GRB10
GC
B
cells
LCL

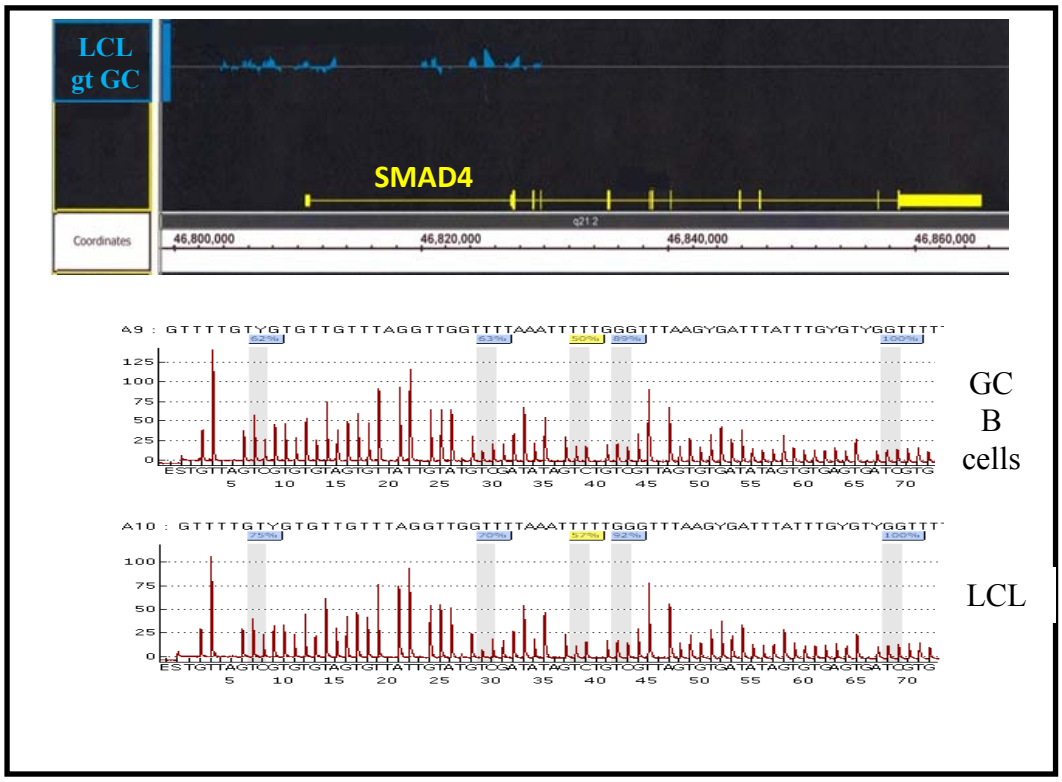
## TCL6


LCL
gt GC
TCL6
Coordinates
95,190,000
95,200,000
95,210,000
GC B
cells
LCL

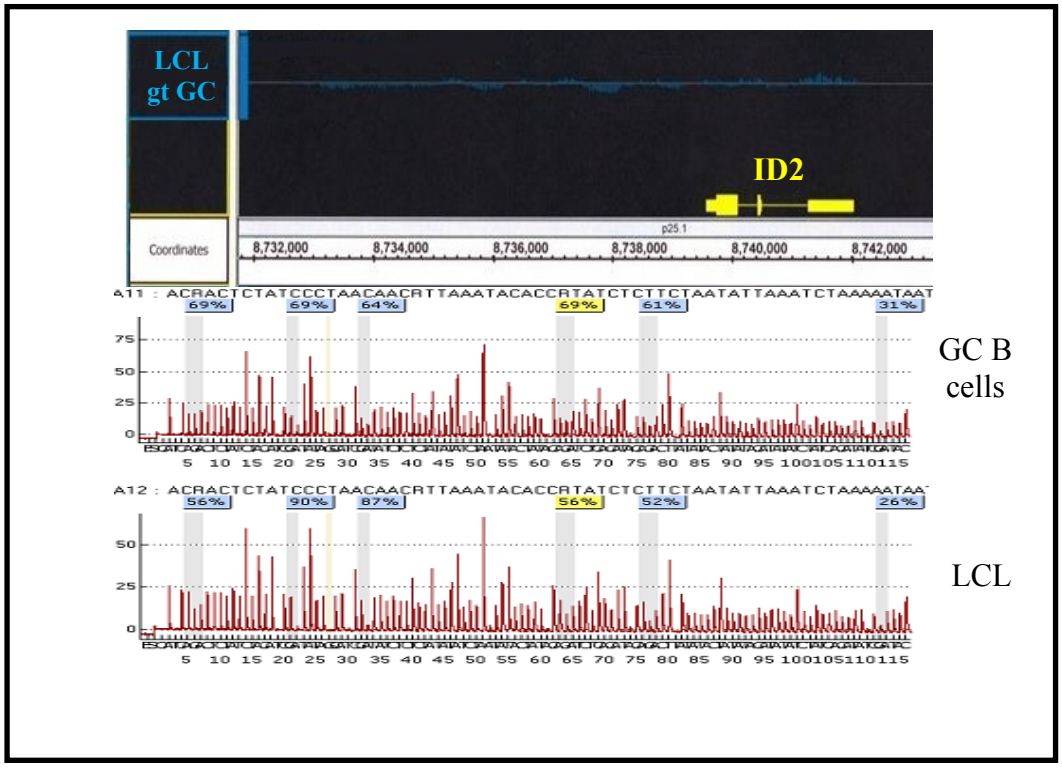
## RBM5


LCL
gt GC
RBM5
Coordinates
50,090,000
50,100,000
50,110,000
50,120,000
50,130,000
GC B
cells
LCL

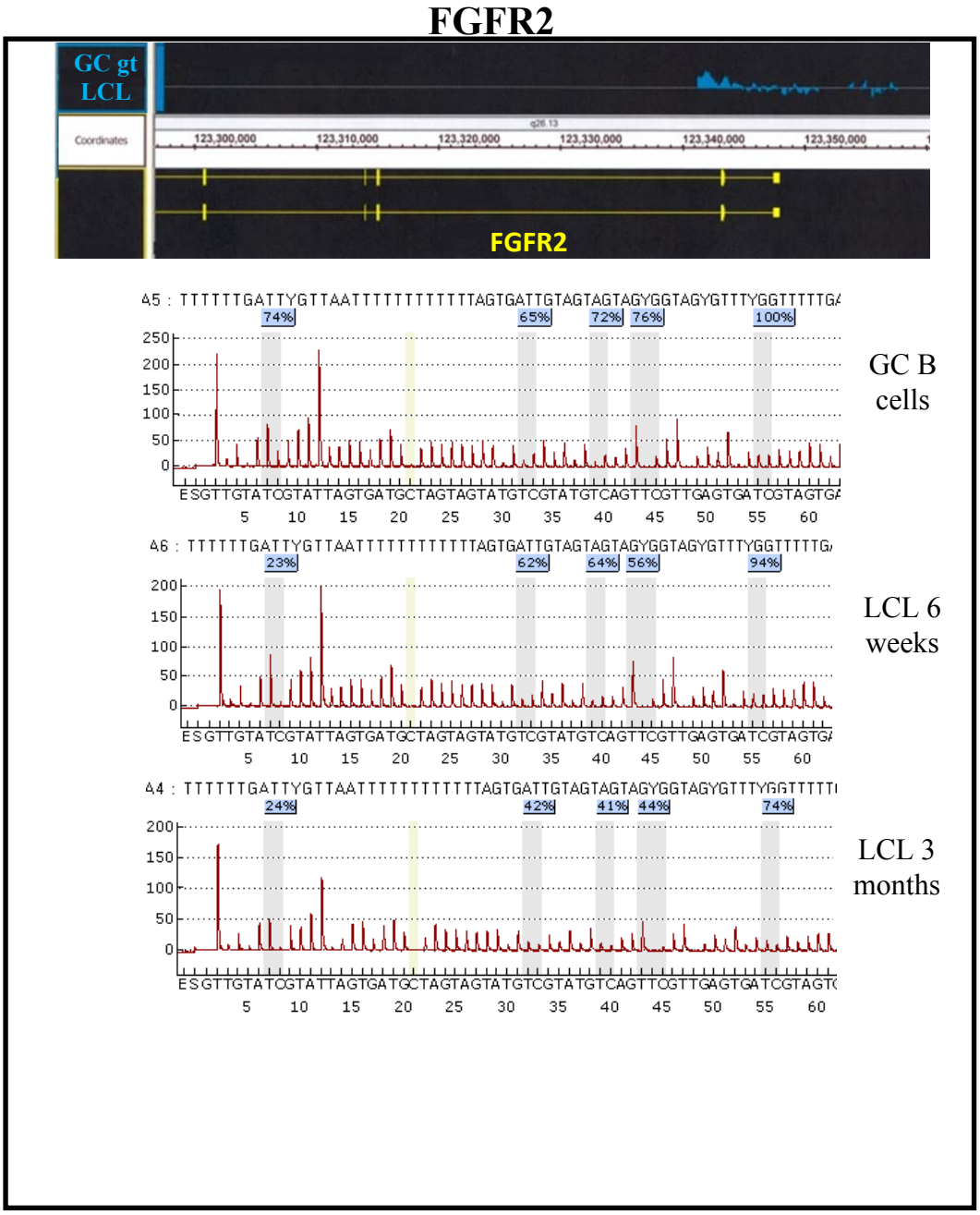
## SMAD4


LCL
gt GC
SMAD4
Coordinates
46,800,000
46,820,000
46,840,000
46,860,000
GC
B
cells
LCL

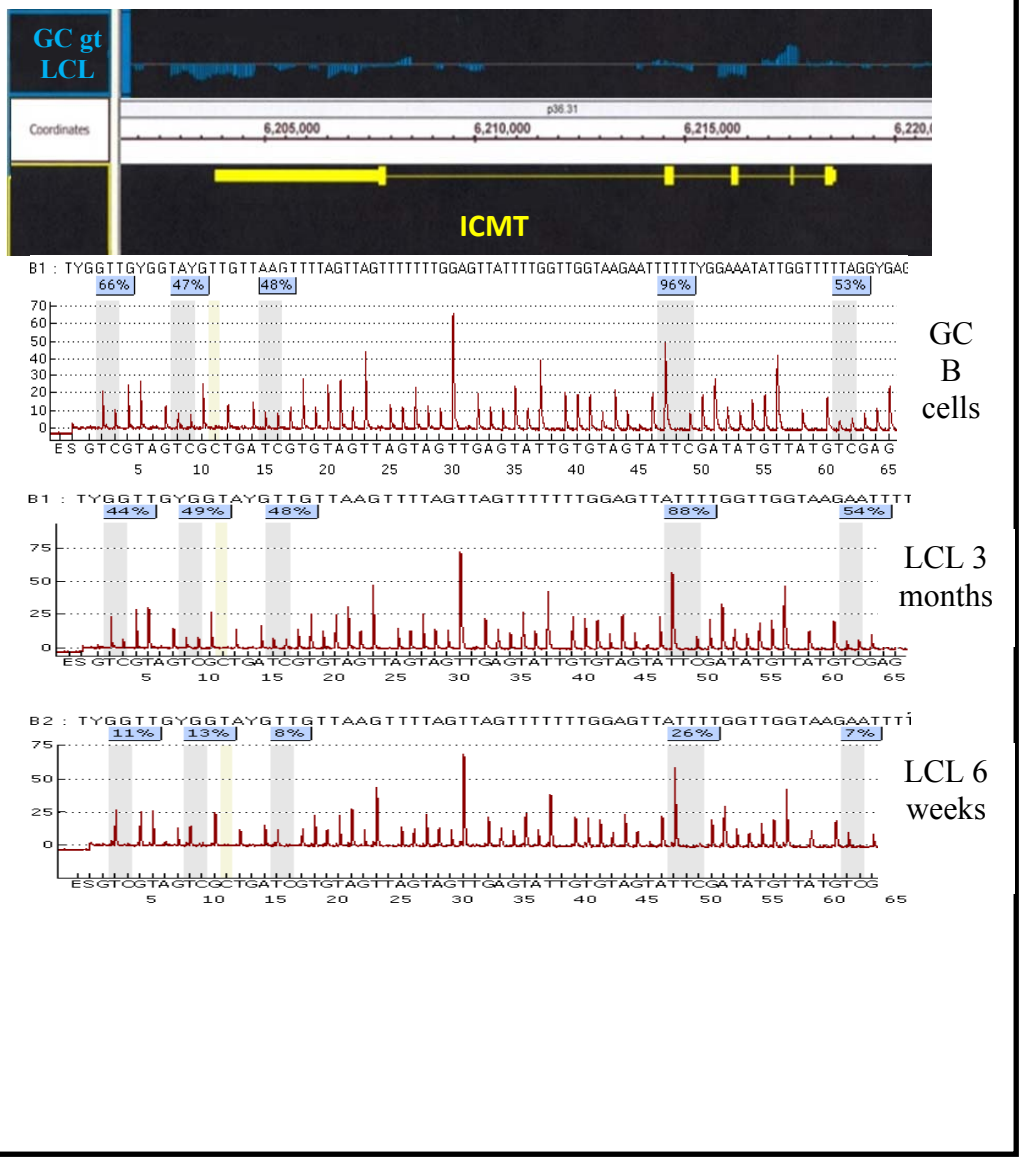
## ID2


LCL
gt GC
ID2
Coordinates
8,732,000
8,734,000
8,736,000
8,738,000
8,740,000
8,742,000
GC B
cells
LCL

Annex 2: FGFR2 and ICMT methylation decreases over time. Pyrosequencing results of FGFR2 and ICMT confirmed the decrease in methylation predicted by the promoter methylation array (6 weeks post infection). Pyrosequencing of these same genes in a 3 month LCL revealed a further derease in methylation.



# ICMT


GC gt LCL
Coordinates
6,205,000
6,210,000
6,215,000
ICMT
GC B cells
LCL 3 months
LCL 6 weeks